

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 7.40034 Seconds

(without alignments)
2028.884 Million cell updates/sec

Title: US-09-697-863A-2

Perfect score: 1902

Sequence: 1 MEIGSCLEGGREAAEEBGP.....DCGRFSDHMLCNDIIL 362

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	140	7.4	429	1 NOCT_MOUSE	035710 mus musculus
2	134.5	7.1	388	1 NOCT_XENLA	P79942 xenopus lae
3	128	6.7	431	1 NOCT_HUMAN	Q9UK39 homo sapien
4	110.5	5.8	361	1 APEA_DICDI	P51173 dictyostell
5	108	5.7	477	1 ISCI_YEAST	P40015 saccharomyc
6	105.5	5.5	253	1 YBHP_ECOLI	P75772 escherichia
7	105.5	5.5	670	1 Y759_HUMAN	Q9UNK9 homo sapien
8	103	5.4	548	1 G6P1_YERPE	Q8ZAS2 yersinia pe
9	102.5	5.4	267	1 EX3_HAEIN	P44318 haemophilus
10	102.5	5.4	549	1 G6P1_SALTI	Q8ZIU7 salmonella
11	102.5	5.4	549	1 G6P1_SALTI	Q8ZIU7 salmonella
12	102	5.4	1391	1 RPOB_MYCPN	Q8ZK14 salmonella
13	99.5	5.2	333	1 PHU1_BACCE	P09599 bacillus ce
14	98.5	5.2	549	1 G6P1_HAEIN	P78013 mycoplasma
15	98.5	5.2	336	1 PGCV_HUMAN	P13611 haemophilus
16	98	5.2	253	1 NOCT_RAT	P13611 homo sapien
17	97.5	5.1	419	1 NSMA_MOUSE	Q9E555 ratius norv
18	97.5	5.1	422	1 NSMA_MOUSE	Q9E555 ratius norv
19	96.5	5.1	422	1 NSMA_MOUSE	Q9E555 ratius norv
20	96.5	5.1	422	1 NSMA_MOUSE	Q9E555 ratius norv
21	93.5	4.9	295	1 YCXY_EUGGR	P31920 euglena gra
22	93.5	4.9	856	1 ENV_HVILMN	P05677 human immun
23	93	4.9	549	1 G6P1_ECOLI	P11537 escherichia
24	93	4.9	1259	1 LINI_HUMAN	P08547 homo sapien
25	92.5	4.9	556	1 PHL_IPIIN	P17627 leptospira
26	92.5	4.9	565	1 NPRI_CANFA	Q9ULI9 canis famli
27	91.5	4.8	807	1 MCM3_XENLA	Q9ULI9 xenopus lae
28	90.5	4.8	193	1 ILI8_HUMAN	Q14116 homo sapien
29	90.5	4.8	657	1 HUTH_HUMAN	P42357 homo sapien
30	90	4.7	507	1 RPP_MEASE	P03422 mesleis vir
31	89.5	4.7	849	1 VNCB_AEDVY	P27454 aedes demo
32	89.5	4.7	587	1 YN4_YEAST	P53933 saccharomyc
33	89.5	4.7	2663	1 CENE_HUMAN	Q02224 homo sapien

34	89	4.7	246	1 YSM4_CAEEL	Q10019 caenorhabdi
35	89	4.7	746	1 STM2_HUMAN	Q9P246 homo sapien
36	88.5	4.7	498	1 FLID_BACSU	P39738 bacillus su
37	88.5	4.7	657	1 HUTH_MOUSE	P35492 mus musculu
38	88.5	4.7	2290	1 POLG_EMCV	P03304 encephalomy
39	88	4.6	387	1 GMD5_DROME	Q9YMW9 drosophila
40	88	4.6	862	1 PGCV_MACNE	Q28858 macaca neme
41	88	4.6	1125	1 MAP4_MOUSE	P27546 mus musculu
42	88	4.6	3678	1 DMD_MOUSE	P11531 mus musculu
43	87.5	4.6	408	1 PKG_SITSO	P50317 sulfolobus
44	87.5	4.6	462	1 YMA5_YEAST	Q04235 saccharomyc
45	87.5	4.6	558	1 G6P1_CRIGR	P50309 cricetus

ALIGNMENTS

RESULT 1
ID NOCT_MOUSE STANDARD; PRT; 429 AA.
AC 035710; Q9QZG9;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Nocturnin (CCR4 protein homolog).
GN CCRN4L OR NOC OR CCR4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=99453012; PubMed=10521507;
RA Dupressoir A., Barbot W., Loireau M.P., Heidmann T.;
RT "Characterization of a mammalian gene related to the yeast CCR4
RT general transcription factor and revealed by transposon insertion.";
RT J. Biol. Chem. 274:31068-31075(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Retina;
RA Wang Y., Osterbur D.L., Green C.B., Besharse J.C.;
RX "Mammalian homologs of Xenopus nocturnin: conservation of structure
RT and circadian regulation.";
RT Submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 62-429 FROM N.A.
RC STRAIN=DBA/2J; TISSUE=Liver;
RX MEDLINE=97190339; PubMed=9038221;
RA Puech A., Dupressoir A., Loireau M.P., Mattei M.G., Heidmann T.;
RT "Characterization of two age-induced intracisternal A-particle-related
RT transcripts in the mouse liver. Transcriptional read-through into an
RT open reading frame-with-similarities to the yeast ccr4 transcription
RT factor.";
RT J. Biol. Chem. 272:5995-6003(1997).
- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR
OR CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH
MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT
LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES (BY SIMILARITY).
- SIMILARITY: BELONGS TO THE CCR4/NOCTURIN FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

EMBL, AF183960; AAD56547.1; -
EMBL, AF199491; AAG01384.1; -
EMBL, U70139; AAB62717.1; ALT_FRAME.
DR MGI:109382; Ccr4.
DR


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF183961; AAD56548.1; -
DR EMBL; AF199492; AAG01387.1; -
DR EMBL; AF199493; AAG01388.1; -
DR EMBL; AF199494; AAG01389.1; -
DR Genew; HGNC:14254; CCRN4L;
DR InterPro; IPR005135; Exo_endo_phos;
DR Pfam; PF03372; Exo_endo_phos; 3.
KW Biological rhythms.
FT CONFLICT 69 T -> N (IN REF. 2).
FT CONFLICT 77 G -> A (IN REF. 2).
FT CONFLICT 266 A -> T (IN REF. 2; AAG01389).
FT CONFLICT 341 341 N -> S (IN REF. 2).
SQ SEQUENCE 431 AA; 48150 MW; B61EF484E8D29A85 CRC64;

Query Match 6.7%; Score 128; DB 1; Length 431;
Best Local Similarity 20.4%; Pred. No. 0.0057;
Matches 86; Conservative 56; Mismatches 156; Indels 124; Gaps 20;

OY 20 PEVKKRRLICVFEFA--SVASCDAAVAOCFLAENDWE-----MERALNSYFEPVEESALE 72
DB 39 PRPASRLIAAASASAGARSCRTV--CSMGTSRLYSGLATLNS-----SAAS 88
OY 73 RRPETISEP-----KTYVDLTNEETDSTSKISPS 103
DB 89 QHPEYLVSPDPEHLPEIPKELLECRALVHTRRPFRQDRVDVDRTD--CPSTHPI--- 143
OY 104 EDTQOENGSMFLITWNI-----DGLD-----LNLSEARAGVSYALSPVIF 149
DB 144 -----RVOMNIIAOLGEGKDNFVOCPEALKMERKCLIEELLAVOPDILC 192
OY 150 LOEVIPPYYSYLKRRSSNYEITGHEGYF-----TAIMLKSRVKL 191
DB 193 LOEV-DHYEDTQPLSLR---LGYQGTFFPKPWSPCIDVEHNNGPDGCLFLQNRFKL 247
OY 192 KSEIIPFSTKMMRNILCV---HVNVSQNELCMTSLHSTEGHAERNNOLKWLK 247
DB 248 VNSANRLTAMTLKTOYALIAOTLECKESGRFCIAVTHLARTGWERFSAQCDLLON 307
OY 248 MOEAPESATV--IFAGDTNLR-DREVTRCGLPN-NIVDVEFL--GKPKHCOTWDTQ 300
DB 308 LQITGAKIPLIVCGDFNAEPTVEYKHFASNSNLNSAVKLLSADQSEPPYTTW--- 364
OY 301 MNSNLGITAAKLRDRIFFRAAAEFGH IPRSDLL-----GLEKIDCGFPDHWGL 355
DB 365 ---KITSGECHRDLTYWYSKHALN---VNSALDLLIEOIGPRLPSFNPYPSDHLSTV 418
OY 356 CN 357
DB 419 CD 420

RESULT 4
APEA.DICDI STANDARD: PRT; 361 AA.
AC P51173;
DT 01-OCT-1996 (Rel. 34; Created)
DT 01-OCT-1996 (Rel. 34; Last sequence update)
DT 15-JUN-2002 (Rel. 41; Last annotation update)
DE DNA (apurinic or apyrimidinic site) lyase (EC 4.2.99.18) (Class II
DE apurinic/apyrimidinic(Ap)-endonuclease).
GN APEA.
OS Dictyostelium discoideum (slime mold).
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostelium.
OX NCBI_TaxId=44689;

```

```

RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96226184; PubMed=8657579;
RA Frieled T.M., Guyer R.B., Ling A.Z., Deering R.A.;
RT "Apurinic/apyrimidinic (Ap) endonuclease from Dictyostelium
RT discoideum: cloning, nucleotide sequence and induction by sublethal
RT levels of DNA damaging agents."
RL Nucleic Acids Res. 24:1950-1953(1996).
CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
CC leaving a 3'-terminal unsaturated sugar and a product with a
CC terminal 5'-phosphate.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC MNNG, GAMMA RAYS, BLEOMYCIN AND STREPTOZOTOCIN
CC -1- INDUCTION: BY DNA-DAMAGING AGENTS INCLUDING ULTRAVIOLET LIGHT,
CC -1- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U31631; AAC47024.1; -
DR HSSP; P27695; 1HD7.
DR Dictydb; DD05067; apea.
DR InterPro; IPR000097; Apendonclse1.
DR InterPro; IPR004442; Exodnase_III.
DR InterPro; IPR004808; ExoIII_xth.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR TIGRFAMS; TIGR00633; xth; 1.
DR TIGRFAMS; TIGR00726; AP_NUCLEASE_FL_1; 1.
DR PROSITE; PS00727; AP_NUCLEASE_FL_2; 1.
DR PROSITE; PS00728; AP_NUCLEASE_FL_3; 1.
KW DNA Repair; Lyase; Nuclear protein.
FT DOMAIN 27 39 POLY-GLU.
FT DOMAIN 41 44 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT METAL 70 80 POLY-GLU.
FT METAL 139 139 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT ACT SITE 351 351 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 361 AA; 41251 MW; 9C4446A0B63641BC CRC64;

Query Match 5.8%; Score 110.5; DB 1; Length 361;
Best Local Similarity 21.5%; Pred. No. 0.12;
Matches 62; Conservative 52; Mismatches 94; Indels 81; Gaps 15;

OY 8 EGGREAEEREPVKKRRRLICVFEFASVASCDAVAOCFLAENMWEMERALNSYFEPVE 67
DB 27 EEEKEVEEED--EEDEKRLVAKKTPAKKA-----PAK 57
OY 68 ESALERRPPTISEPRYYVDLTNEETDSTSKISPSEDTOE---NCSMFLTWINDGL 124
DB 58 KAAAKKRSK--DEDEDEEKEEETNKTTASVSIAIDNDEPVEENOMKIISWNAIG 115
OY 125 DLNLSERARGVSYALYSPVIFLOEV-IPPYYSYTK---RSSNYEITGHEGYF 179
DB 116 K---SVLSKGTCEVEKENPDVCLQETKTP--SNIKKQOMKRGYEYHIEDQGGHH 169
OY 180 -TAIMLKSRVRLKSOELIPFSTKMMRNILCVHNVSGNELC-----MTSHLES-- 229
DB 170 GGVGLTKKK-----PNAITFGIGIAKHON--EGRVITLEYVOFYIVNYPINAG 216
OY 230 TEGHAERNNOLKWLK-----MOEAPESATVIFAGDTNLRDEYV 270
DB 217 TRG-----LQRLDYRIKEDWDFOAVLEKLNATKFTLWCGDLNAHTET 260

RESULT 5
ISCL_YEAST

```

```

ID ISCL1_YEAST STANDARD: PRT: 477 AA.
AC P40015;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Inositol phospholipase C (EC 3.1.4.-) (IPS
DE phospholipase C) (IPS-PLC) (Neutral sphingomyelinase) (N-Smase)
DE (nsmase).
GN ISCL1 OR YER019W.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycas.
ON NCBI_TaxID=4932;
RX MEDLINE=20564359; PubMed=11006294;
RA Hannun Y.A.;
RA "Identification of ISCL1 (YER019W) as inositol phospholipase C in
RA phospholipase C in Saccharomyces cerevisiae.";
RT J. Biol. Chem. 275:39793-39798(2000).
CC -1- FUNCTION: Responsible for the hydrolysis of the
CC phospholipids (IPs), inositol phosphorylceramide (IPC),
CC mannosylinositol phosphorylceramide (MIPC), and mannosylinositol
CC phosphorylceramide (M(IP)2C). Also active on sphingomyelin, but
CC this activity is probably not physiologically relevant.
CC -1- COPACITOR: Magnesium.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
CC -1- SIMILARITY: BELONGS TO THE NEUTRAL SPHINGOMYELINASE FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U18778; AAB64552.1; -.
CC SGP: S0000821; ISCL1.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos.1.
KW Hydrolase; Magnesium; Transmembrane.
FT TRANSMEM 399 417 POTENTIAL.
FT TRANSMEM 425 449 POTENTIAL.
FT METAL 100 100 MAGNESIUM (BY SIMILARITY).
FT SITE 233 233 IMPORTANT FOR SUBSTRATE RECOGNITION (BY
FT SIMILARITY).
FT ACT_SITE 334 334 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 477 AA; 53940 MW; 0670FD303FEB8EFF CRC64;
Query Match 5.7%; Score 108; DB 1; Length 477;
Best Local Similarity 22.0%; Pred. No. 0.28;
Matches 78; Conservative 47; Mismatches 124; Indels 106; Gaps 20;
QY 90 EETDTSTSKSPSEDPTQENGSMFSL--ITWINDGLDL--NNLSERRGCVSYALAYSP 145
DB 13 KEDGSEFALNGTNAIMNSDKAVSIRKLFNTWGLKVVSKRKRRLRAIADKLASHSM 72
QY 146 -----DVIFLOWY-IPPYVSYLKKR--SSNYEILITGHEEGYFT-- 180
DB 73 LRPISDELLPNCGSDSNENEDYVIALQETWCEVDKWKIYACASAKYPPYQRLFHSSILTP 132

```

```

QY 181 --AHLKRSRYLKSQELIIPP-----STKMMRLNCYHVNSGNELCMTSHL 227
DB 133 GLAIL---SKVIESTFLYRPIINGRPSAVFRGMYVKSLAIVLTGTPPIAIMNSHM 189
QY 228 E---SNRGHAERMQ-----LKMVLKMKQEPASATVIFAGDTNLRDREYTRCGGLPN 278
DB 190 HAPYAKGDAALCHRSQCAMDFSLIKLYRQA--GYAVIVGDLN-----SRGSLPH 241
QY 279 -----NIVDWEFL-GKP-----KHQ-----YTMDIQMNSLGLTA 309
DB 242 KFLTGAGLVDSMEDLHGKODLAVIARLSPLQQLKGCTGDSILNTWRARQDP---E 297
QY 310 ACKLRP---DRIFFRMAAEHGHIIPRSIDLGLFEKDCGRFSDHGLCNLDII 361
DB 298 ACRDLIALIDPPLQTVLDVAGVFTER-----IPHLDS--VSDHFAVSCLTNIY 344
RESULT 6
YBHP_ECOLI STANDARD: PRT: 253 AA.
ID YBHP_ECOLI
AC P75772;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein ybhp.
GN YBHP OR B0790 OR Z1009 OR ECS0868.
OS Escherichia coli.
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
ON NCBI_TaxID=562, 83334;
RX MEDLINE=97426617; PubMed=9278503;
RA Blatner F.R., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN
RN SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sempel G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-Kb DNA sequence of the Escherichia coli K-12 genome
RT corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDU933 / ATCC 700927;
RX MEDLINE=21074935; PubMed=11206551;
RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dinalanta E.T., Potamoudis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blatner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL Nature 409:529-533(2001).
RN
RN SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 0509952;
RX MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,

```



```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-CO-92 / Biovar Orientalis;
RX      MEDLINE-21470413; PubMed-11586360;
RA      Parkhill J., Wren B.W., Thomson N.R., Titchall R.W., Holden M.T.G.,
RA      Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA      Baker S., Basham D., Bentley S.D., Brooks K., Cerdono-Tarrasa A.M.,
RA      Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA      Felwell T., Hamlin N., Holroyd S., Jagsels K., Karyshev A.V.,
RA      Leather S., Moule S., Oyston P.C.F., Quail M., Ratheford K.,
RA      Simmonds M., Skelton J., Stevens K., Whitehead S., Barrett B.G.;
RT      "Genome sequence of Yersinia pestis, the causative agent of plague.";
RL      Nature 413:523-527(2001).
CC      -1- CATALYTIC ACTIVITY: D-glucose 6-phosphate = D-fructose 6-
CC      phosphate.
CC      -1- PATHWAY: Involved in glycolysis and in gluconeogenesis.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC      -1- SIMILARITY: BELONGS TO THE GPI FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      -----
DR      EMBL; AJ414158; CAC93186.1; -.
DR      InterPro; IPR001672; G6P_Isomerase.
DR      Pfam; PF00342; Pgi: 1.
DR      Pfam; PRO0662; G6PISOMERASE
DR      PROSITE; PS00765; P_GLYCOCSE_ISOMERASE_1; 1.
DR      PROSITE; PS00174; P_GLYCOCSE_ISOMERASE_2; 1.
KW      Isomerase; Gluconeogenesis; Glycolysis; Complete proteome.
FT      ACT_SITE 386 386 By similarity.
FT      ACT_SITE 514 514 By similarity.
SQ      SEQUENCE 548 AA: 61161 MW: 3ED9D9B3A308A3B CRC64:

Query Match 5.4%; Score 103; DB 1; Length 548;
Best Local Similarity 21.1%; Pred. No. 0.85;
Matches 75; Conservative 57; Mismatches 106; Indels 118; Gaps 20;

QY 56 RALNATYFEPPEVESALERRPETIS-----EPKYYVDLTNETTDSITS 98
DB 13 KALQOHFE-----QMKDVTISLFAKDDQRFNRFSAFPDDQMLDYFSKNRITSETLE 64
QY 99 KISPSEDDQNG-----SMFSLITWNIDGLDNLNLSRARGVCSYAL----YSPDYI 148
DB 65 KL--QDLAKETDLGAIKSMFS-----GKIKNTEDRA--VLIATLRNRSNTPIYV 111
QY 149 FLOEYVIPYYSYLKRSSNYE-IINGHEGYFTATMLKSRYLKSOELIPEPSTRMMR- 206
DB 112 DGDVMPVEPNVNLAMKQFCQDVRISGDMKGYGKAITVDVNNIGIGSDGCPMYEALRP 171
QY 207 --NLICVHVNSGNELCIMTSHLESTRGHAERMNOLK-----VLLKMOEAPESAT- 256
DB 172 YKNHLNMH-----FVSNVDTG--HIEALPLNPETTLFLVASRTFTQETMTNA 219
QY 257 -----YIFAGD-----TNLRDREYTCGGGLPNIVDWEFLGPKPKCYOT- 296
DB 220 HSARDWFLSAADDPAHVAKHAFAALSTNA--KAVGEFGIDTNMFFWVGVS---RVS 273
QY 297 WDTQNSNLTGTAACKLRF-----DRIFFAAAEGHIIPRSLDLGL 339
DB 274 W-----SAIGLSIALSVGEFHEPQLLSGAHADKRFALFPAEKXN--LVLALLIGI 322

RESULT 9
EX3_HAEIN
ID EX3_HAEIN STANDARD: PRT: 267 AA.
AC P4A318:
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)

```

```

DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Exodeoxyribonuclease III (EC 3.1.11.2) (Exonuclease III) (EXO III).
OS XTRA OR H10041.
OC Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kinkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McInney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spilligs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Elie L.D., Fitchman J.L., Fuhrmann J.L., Georgagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd."
RL Science 269:496-512(1995).
CC -I- FUNCTION: MAJOR APURINIC-APYRIMIDINIC ENDONUCLEASE OF E. COLI. IT
CC REMOVES THE DAMAGED DNA AT CYTOSINES AND GUANINES BY CLEAVING ON
CC THE 3' SIDE OF THE AP SITE BY A BETA-ELIMINATION REACTION
CC (BY SIMILARITY).
CC -I- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
CC progressively in a 3'-to 5'-direction, releasing nucleoside 5'-
CC phosphates.
CC -I- SUBUNIT: MONOMER (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32689; AAC21719.1; -.
DR HSSP: P09030; IAKO.
DR TIGR: H10041; -.
DR InterPro: IPR000097; Apendonclse1.
DR InterPro: IPR004442; ExoDNase_III.
DR InterPro: IPR004808; ExoIII_xth.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR TIGRFAMS: TIGR00195; exodnase_III; 1.
DR TIGRFAMS: TIGR00633; xth; 1.
DR PROSITE: PS00726; AP_NUCLEASE_F1_1; 1.
DR PROSITE: PS00727; AP_NUCLEASE_F1_2; 1.
DR PROSITE: PS00728; AP_NUCLEASE_F1_3; 1.
DR Hydrolase; Nuclease; Exonuclease; DNA repair; Complete proteome.
KW METAL 34
FT ACT_SITE 259 259 MACENESUM OR MANGANESE (BY SIMILARITY).
FT ACET_SITE 259 259 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 267 AA; 31025 MW; 6B3ADE465A1E347C CRC64;

Query Match 5.4%; Score 102.5; DB 1; Length 267;
Best Local Similarity 21.1%; Pctd. No. 0.34;
Matches 59; Conservative 45; Mismatches 108; Indels 67; Gaps 13;

QY 117 ITWNIDGDLNNLSRRAGVCSYALATSPVIFLOE--VIPPYSYLKRKSSNYELITGH 174
Db 4 ISFNING-----LARRPHOLEAITEKQPDVITGLQETKVADEAFPEITENLGYHFHNG 58
QY 175 EGGYETALMLKRSRYKLQSEILIPFPSTKMRRLLCVHVNVSQNELCLMT----- 224
Db 59 QKGHGVALLIKQPEK-IRRGFPTDNEDQKRIMADE---TEGGLTVINGVFPGE 114
QY 225 SHLESTRGHAERK-NOLKMYLKMQDAPSAVIYIFAGDTNLRDEVT----- 271

```

QY 85 VDLTNEETDSTSKISPSDEDTQQENG-----SMFSLITWNIDGLDLNNLSERAGCVCS 138
|| : | : || : : ||| | : : ||
Db 51 VDFSKNRTEETLAKL--ODLAKETDLGAIKSMFS-----GEKINRTEDRA---YL 97

Protein	Enzyme	EC	Reaction
DR	PROSITE: PS00765; P-GLUCOSE-ISOMERASE_1;	1.	
DR	PROSITE: PS00174; P-GLUCOSE-ISOMERASE_2;	1.	
KW	Isomerase: Glucoseisomerases; Glycolysis; Complete proteome		
FT	ACT_SITE	386	BY SIMILARITY.
FT	ACT_SITE	514	BY SIMILARITY.

```

SQ SEQUENCE 549 AA: 61428 MW: 85642951920C4E6C CRC64:
Query Match 5.4%; Score 102.5; DB 1; Length 549;
Best Local Similarity 21.3%; Pred. No. 0.94;
Matches 67; Conservative 55; Mismatches 90; Indels 103; Gaps 18;

QY 85 VDLTNEETDSTSTKISPSDFTQOENG-----SMFSLTNNIDGLDNLNLSERARGVCS 138
   || : || : || : || : || : || : || : || : || : || : || : || : ||
DB 51 VDFSKRRTTEETLAKL---QDLAKETDLGAIKSMFS-----GEKINRTDRA---VL 97

QY 139 YIAL---YSPDVIFLOEVIIPPYSYLKK-RSSNYEITGHEGYFTAIMLKSRVYKLS 193
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 98 HVALNRSMPTIIVDGKDVMEVNAVLEKMKTFSSAIIISGQKGTGKAITDVNVIIGG 157

QY 194 QETIPRPSRKMR---NLLCVHNVNSGNELCMTSHLESTRGHAEKRNQMKVYKMOE 250
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 158 SDGPEPMVEALRPYKNHLMH-----FVSNVDGT-HIAE-----VLKYN- 197

QY 251 APESATVIFAGDT-----NLRD-----REVTRCGGL 276
   || : || : || : || : || : || : || : || : || : || : || : ||
DB 198 -PETITFLVASKTFTTQETMTNAHSARDWFLKTAGEDEKHVAKHFAALSTNAKAVEFGID 256

QY 277 PNNIVDWEFLGKPKHCQYTT-WDTQMSNLGITTAACKLRFD--RIFPRAAECHT--- 329
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 257 TANMEFFEMVVGQ---RYSLW-----SAIGLSILSVGFDNFVELLSGAHAMDKHFSYT 307

QY 330 -----IPRSLDLGL 339
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 308 PAKKNLPILLALIGI 322

RESULT 12
PROB_MYCPN STANDARD; PRT; 1391 AA.
ID PROB_MYCPN STANDARD; PRT; 1391 AA.
AC F78013;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE DNA-directed RNA polymerase beta chain (EC 2.7.7.6) (Transcriptase
DE beta chain) (RNA polymerase beta subunit).
GN RPOB OR MPM516 OR MP526.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97105885; PubMed=8948633;
RA Himmelfeich R., Hilbert H., Plagens H., Pirkel E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RL Nucleic Acids Res. 24:4420-4449 (1996).
CC -I- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
CC OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA)(N).
CC -I- SUBUNIT: THE ENZYME CONSISTS OF THE SIGMA CHAIN AND THE CORE
CC ENZYME WHICH IS COMPOSED OF 2 ALPHA CHAINS, 1 BETA CHAIN, AND 1
CC BETA' CHAIN.
CC -I- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA CHAIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE000030; AAB95974.1; -.
CC HSSP: O9KM07; IHOM.

```

```

DR InterPro: IPR001572; RNA_POL_B.
DR Pfam: PF00562; RNA_POL_B.1
DR PROSITE: PS01166; RNA_POL_BETA; 1.
DR Transferrase: Transcription; DNA-directed RNA polymerase;
RW Complete proteome.
SQ SEQUENCE 1391 AA: 155621 MW: B2F345AB24F18EAD CRC64:
Query Match 5.4%; Score 102; DB 1; Length 1391;
Best Local Similarity 22.1%; Pred. No. 3.8; Indels 120; Gaps 24;
Matches 85; Conservative 64; Mismatches 90; Indels 120; Gaps 24;

QY 44 QCEFLAENDWEMERALNSFFEPPEV-----ESALRR-RPETISEPKT----- 83
   || : || : || : || : || : || : || : || : || : || : || : ||
DB 42 QTFLL---DHDELNLIAVLF--PIKSPNDRTYINFKGLRTPAPERNEAQSSEKTYEIGI 96

QY 84 YVDLTNEETDSTSTKISPSDFTQOENGSMFSLTNNIDGLDNLNLSERARGVCSIALY 143
   || : || : || : || : || : || : || : || : || : || : || : ||
DB 97 YADL---ELIDSATGATIKPKRSKNSA-----TSSVDGVFLTN-----LPLI 136

QY 144 SPDVIFLOEVI-----PPYV-----SYLKRSSNYEITGHEGYFT-----A 181
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 137 TRDGVFTYINGIEKRVIAQITRSPGIIYMLTSQKLSSRRV---QEGYVCEVLNANGS 192

QY 182 IML-----KSRVYKLSQEIIT-----PFPSTKMRNLICVHNVNSGNELC----- 221
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 193 VMLIYISNKKKIIEADAFVQILLRDVAREGAKIFPITLLKAF-----GMSGKEILKVFKN 247

QY 222 -LMTSHLESTRGHAEKRN-----QMKVYKMOEAPPSAIVITAGDNLNLRDREVTRCGGL 276
   || : || : || : || : || : || : || : || : || : || : || : ||
DB 248 EFTIRSLAEVYNNAKKDELNVNDEPKIKMLREFRD-----GKTDLRKKIADSDKI 297

QY 277 PNNIVDWEFLGKPKHCQYTT-WDTQMSNLGITTAACKLRFDRIFPRAAECHTIFR-SL 334
   : || : || : || : || : || : || : || : || : || : || : || : ||
DB 298 RSLVSDY--VLLKEHKAHLSKARNDEPKVQLEADMDLMDKITTEHAAK--HIVHLSI 353

QY 335 DLGLERKLD-CGRPPSDHWGLCN 357
   || || || || : || : || : || : || : || : || : || : || : ||
DB 354 SLRGLNTEDECP--NSYHALICS 375

RESULT 13
PHL1_BACCE STANDARD; PRT; 333 AA.
ID PHL1_BACCE STANDARD; PRT; 333 AA.
AC P09599;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Sphingomyelinase C precursor (EC 3.1.4.12) (Sphingomyelin
DE phosphodiesterase) (Smase) (SMPIC) (Cereolysin B).
GN Sph.
OS Bacillus cereus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1396;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SE-1;
RX MEDLINE=89057484; PubMed=2848222;
RA Johansen T., Haugli F.B., Ikezawa H., Little C.;
RT "Bacillus cereus strain SE-1: nucleotide sequence of the
RT sphingomyelinase C gene.";
RL Nucleic Acids Res. 16:10370-10370 (1988).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=VKM 164;
RX MEDLINE=93249510; PubMed=8387306;
RA Kuzmin N.P., Gavrilenko I.V., Krukov V.M., Karpov A.V.;
RT "Nucleotide sequence of phospholipase C and sphingomyelinase genes
RT from Bacillus cereus BKM-B164 (letter).";
RL Bioorg. Khim. 19:133-138 (1993).
RN [3]
RP SEQUENCE OF 1-188 FROM N.A.
RC STRAIN=SE-1;
RX MEDLINE=88313678; PubMed=3137122;

```


AC P13611; P20754; Q90NM5; Q13010; Q13189; Q15123;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Versican core protein precursor (large fibroblast proteoglycan)
 DE (Chondroitin sulfate proteoglycan core protein 2) (PG-M) (Glial
 DE hyaluronate-binding protein) (GHAHP).
 GN CSFG2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A. (ISOFORM V0).
 RX MEDLINE=95105188; PubMed=7528742;
 RA Naso M.F., Zimmermann D.R., Iozzo R.V.;
 RT "Characterization of the complete genomic structure of the human
 RT versican gene and functional analysis of its promoter.";
 RL J. Biol. Chem. 269:32999-33008(1994).
 RN [2]
 RP SEQUENCE FROM N.A. (ISOFORM V1).
 RC TISSUE=Placenta;
 RX MEDLINE=90059882; PubMed=2583089;
 RA Zimmermann D.R., Ruoslahti E.;
 RT "Multiple domains of the large fibroblast proteoglycan, versican.";
 RL EMBO J. 8:2975-2981(1989).
 RN [3]
 RP SEQUENCE FROM N.A. (ISOFORM V2).
 RC TISSUE=Glial tumor;
 RX MEDLINE=95105187; PubMed=7806529;
 RA Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "A novel glycosaminoglycan attachment domain identified in two
 RT alternative splice variants of human versican.";
 RL J. Biol. Chem. 269:32992-32998(1994).
 RN [4]
 RP SEQUENCE OF 2711-3396 FROM N.A.
 RC TISSUE=Lung fibroblast;
 RX MEDLINE=88007514; PubMed=2820964;
 RA Krusius T., Gehlsen K.R., Ruoslahti E.;
 RT "A fibroblast chondroitin sulfate proteoglycan core protein contains
 RT lectin-like and growth factor-like sequences";
 RL J. Biol. Chem. 262:13120-13125(1987).
 RN [5]
 RP SEQUENCE OF 251-347 FROM N.A.
 RX MEDLINE=93122792; PubMed=1478664;
 RA Iozzo R.V., Naso M.F., Cannizzaro L.A., Wasmuth J.J.,
 RA McHersson J.D.;
 RT "Mapping of the versican proteoglycan gene (CSPG2) to the long arm of
 RT human chromosome 5 (5q12-5q14).";
 RL Genomics 14:845-851(1992).
 RN [6]
 RP SEQUENCE FROM N.A. (ISOFORM V3).
 RC TISSUE=Brain;
 RX MEDLINE=95181355; PubMed=7876137;
 RA Zako M., Shiomura T., Ujita M., Ito K., Kimata K.;
 RT "Expression of PG-M(V3), an alternatively spliced form of PG-M
 RT without a chondroitin sulfate attachment in region in mouse and human
 RT tissues.";
 RL J. Biol. Chem. 270:3914-3918(1995).
 RN [7]
 RP SEQUENCE OF 3333-3396 FROM N.A. (ISOFORM VINT).
 RC TISSUE=Aortic smooth muscle;
 RX MEDLINE=99327053; PubMed=10397680;
 RA Lemire J.M., Braun K.R., Maurel P., Kaplan E.D., Schwartz S.M.,
 RA Wight T.N.;
 RT "Versican/PG-M isoforms in vascular smooth muscle cells.";
 RL Arterioscler. Thromb. Vasc. Biol. 19:1630-1639(1999).
 RN [8]
 RP PARTIAL SEQUENCE.
 RC TISSUE=Brain;
 RX MEDLINE=89174663; PubMed=246683;
 RA Perides G., Lane W.S., Andrews D., Dahl D., Bignami A.;
 RT "Isolation and partial characterization of a glial

RT hyaluronate-binding protein.";
 RL J. Biol. Chem. 264:5981-5987(1989).
 RN [9]
 RP TISSUE SPECIFICITY OF ISOFORMS.
 RX MEDLINE=96213482; PubMed=8627343;
 RA Paulus W., Baur I., Dours-Zimmermann M.T., Zimmermann D.R.;
 RT "Differential expression of versican isoforms in brain tumors.";
 RL J. Neuropathol. Exp. Neurol. 55:528-533(1996).
 CC -1- FUNCTION: May play a role in intercellular signaling and in
 CC connecting cells with the extracellular matrix. May take part in
 CC the regulation of cell motility, growth and differentiation. Binds
 CC hyaluronic acid.
 CC -1- SUBCELLULAR LOCATION: Secreted; extracellular matrix.
 CC -1- ALTERNATIVE PRODUCTS: At least 5 isoforms: V0 (shown here), V1,
 CC V2, V3 and Vint; are produced by alternative splicing.
 CC -1- TISSUE SPECIFICITY: Cerebral white matter. V0 and V1 is expressed
 CC in normal brain, gliomas, medulloblastomas, schwannomas,
 CC neurofibromas, and meningiomas; V2 is restricted to normal brain
 CC and gliomas; V3 is found in all these tissues except
 CC medulloblastomas.
 CC -1- DEVELOPMENTAL STAGE: Disappears after the cartilage development.
 CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
 CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
 CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
 CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
 CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
 CC -----
 CC *This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-slb.ch/announce/>
 CC or send an email to license@isb-slb.ch).
 CC -----
 DR EMBL: U16306; AAA65018.1; -
 DR EMBL: X15998; CAA34128.1; -
 DR EMBL: S52488; AAB24878.1; -
 DR EMBL: U26555; AAA67565.1; -
 DR EMBL: D32039; BAA06801.1; -
 DR EMBL: J02814; AAA56487.1; -
 DR EMBL: AF084545; AAD48545.1; -
 DR PIR: S06014; S06014.
 DR PIR: A29348; A29348.
 DR PIR: A30358; A30358.
 DR HSSP: P01132; 1EGF.
 DR GeneW: HGNC:2464; CSPG2.
 DR MIM: 118661; -
 DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF-2.
 DR InterPro: IPR001881; EGF-Ca.
 DR InterPro: IPR001438; EGF-II.
 DR InterPro: IPR003599; I9.
 DR InterPro: IPR003006; I9_MHC.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR000538; Link.
 DR InterPro: IPR000436; Sush1_SCR_CCP.
 DR Pfam: PF00008; EGF; 2.
 DR Pfam: PF00047; Ig; 1.
 DR Pfam: PF00059; Lectin_C; 1.
 DR Pfam: PF00084; Sush1; 1.
 DR Pfam: PF00193; Xlink; 2.
 DR PRINTS: PR00010; EGFBIOD.
 DR PRODOM: PD000918; Link; 2.
 DR SMART: SM00032; CCP; 1.
 DR SMART: SM00034; CLECT; 1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00001; EGF_like; 1.
 DR SMART: SM00409; Ig; 1.
 DR SMART: SM00445; Link; 2.
 DR PROSITE: PS00010; ASX_HYDROXYL; 1.

DR PROSITE: PS00022; EGF_1; 2.
DR PROSITE: PS01186; EGF_2; 1.
DR PROSITE: PS01187; EGF_CA; 1.
DR PROSITE: PS01241; LINK; 2.
DR PROSITE: PS00615; C-TYPE LECTIN 1; 1.
DR PROSITE: PS50041; C-TYPE LECTIN 2; 1.
KW Glycoprotein; Proteoglycan; Lectin; Extracellular matrix; Sushi;
KW Signal; Repeat; EGF-like domain; Calcium; Immunoglobulin domain;
KW Hyaluronic acid; Alternative splicing.
FT SIGNAL 1 20 POTENTIAL.
FT CHAIN 21 3396 VERSICAN CORE PROTEIN.
FT DOMAIN 37 137 IG-LIKE V-TYPE DOMAIN.
FT DOMAIN 167 244 LINK 1.
FT DOMAIN 265 346 LINK 2.
FT DOMAIN 348 1335 GAG-ALPHA (GLUCOSAMINOGLYCAN ATTACHMENT
DOMAIN).
FT DOMAIN 1336 3089 GAG-BETA.
FT DOMAIN 3089 3125 EGF-LIKE 1.
FT DOMAIN 3127 3163 EGF-LIKE 2, CALCIUM-BINDING (POTENTIAL).
FT DOMAIN 3176 3290 C-TYPE LECTIN.
FT DOMAIN 3295 3353 SUSHI.
FT DISULFID 44 130 BY SIMILARITY.
FT DISULFID 172 243 BY SIMILARITY.
FT DISULFID 196 217 BY SIMILARITY.
FT DISULFID 270 345 BY SIMILARITY.
FT DISULFID 294 315 BY SIMILARITY.
FT DISULFID 3093 3104 BY SIMILARITY.
FT DISULFID 3098 3113 BY SIMILARITY.
FT DISULFID 3115 3124 BY SIMILARITY.
FT DISULFID 3131 3142 BY SIMILARITY.
FT DISULFID 3136 3151 BY SIMILARITY.
FT DISULFID 3159 3162 BY SIMILARITY.
FT DISULFID 3169 3180 BY SIMILARITY.
FT DISULFID 3197 3289 BY SIMILARITY.
FT DISULFID 3265 3281 BY SIMILARITY.
FT DISULFID 3296 3339 BY SIMILARITY.
FT DISULFID 3325 3352 BY SIMILARITY.
FT CARBOHYD 57 57 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 615 615 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 782 782 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 809 809 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1332 1332 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1398 1398 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1442 1442 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1468 1468 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1663 1663 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 1898 1898 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2179 2179 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2272 2272 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2280 2280 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2360 2360 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 2385 2385 N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 5.2%; Score 98.5; DB 1; Length 3396;
Best Local Similarity 22.0%; Pred. No. 26;

Matches 76; Conservative 53; Mismatches 131; Indels 85; Gaps 18;

QY 7 LEGREAAEEGPEVKKRRLLCVEF-----ASVASCDA--VAQCFLAENDWEMERAL 58
DB 2066 VEGTKAPVEKE---EYKVSSTVSTNFPOTIEPAKLMSROEVNPNVROEISEFTTSEEQIOE 2122
QY 59 NSYFEPPEVSALERRPETISEPKTYVDLTNEETTD---STTSKISPSEDTQOENGSMF 114
DB 2123 EKSFESPONSPATE---QTTFDSQTFTE-TELKTDYSLVTKTKTYSDDKEMKEEDTSLV 2178
QY 115 SLTTWNIDGIDLNNLSERAGVCSYLALSPD-----VIFLOEVIPPIY-----SY 160
DB 2179 NMSTPDDP-----ANGLESYTTL--PEATEKSHFFLATALVATESIPAHEVVTDSP 2226
QY 161 LKRRSNVEIITGHEGYFTAILMKSRVKL-----KSOELIP-----FPSTKMR 206
DB 2227 IKKEESTKHPKMRP-----TIQESDTLELFSGLSGGEVPLPTLESVNFTEVEQIN 2280

QY 207 NLLCVHWVWVSGNELCLMTSHLESTRGHAERBNOLKMYLKKMOEAP-ESATVIFAGDTNL 265
DB 2281 NTLVPH-----TSQVESTSPDKIEDFNMENVAREV--GPLVSOQDIFEGSGSV 2327
QY 266 RDR---EYTRCGGLPNNIYDVWEE---LGKPKHQYTWDMQNSN 304
DB 2328 TSTLLIEILSDPTGAGCPYVADLPSTDIGHPQNOTVRMAEIQTS 2372

Search completed: November 14, 2002, 10:25:45
Job time : 12.4003 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 13.5673 Seconds
(without alignments)
2565.041 Million cell updates/sec

Title: US-09-697-863a-2

Perfect score: 1902

Sequence: 1 MRLGSCLEGGRAAEBCPP.....DCGRPSDHMLCNLDIIL 362

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR-73:*

1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	420	22.1	362	2 T27272	hypothetical prote
2	231	12.1	441	2 B86252	hypothetical prote
3	113	5.9	502	2 T40792	hypothetical prote
4	111.5	5.9	252	2 D95853	conserved hypothet
5	110.5	5.8	361	2 S68268	apurinic/apurimidi
6	108	5.7	253	2 T35506	hypothetical prote
7	108	5.7	477	2 S50477	hypothetical prote
8	105.5	5.5	253	2 D90737	hypothetical prote
9	105.5	5.5	253	2 E85587	hypothetical prote
10	105.5	5.5	253	2 F64815	yhbp protein - Esc
11	105.5	5.5	673	2 T46340	hypothetical prote
12	105	5.5	1374	2 A84888	hypothetical prote
13	104.5	5.5	281	2 AG2333	hypothetical prote
14	103.5	5.4	236	2 AD1115	hypothetical prote
15	103	5.4	288	2 T24066	exonuclease III ho
16	103	5.4	548	2 AF0452	glucose-6-phosphat
17	102.5	5.4	267	2 H64044	exodeoxyribonuclea
18	102.5	5.4	549	2 AD1013	glucose-6-phosphat
19	102	5.4	265	2 F97566	probable nuclease
20	102	5.4	265	2 AD2787	exonuclease III [1
21	101	5.4	1391	2 S73652	RNA polymerase bet
22	101	5.3	1275	2 I38568	reverse transcript
23	101	5.3	1275	2 S65824	reverse transcript
24	100	5.3	1261	2 E59430	PP1L-associated R
25	99.5	5.2	333	2 S01950	sphingomyelin phos
26	99.5	5.2	592	2 P80197	phospholipase C (E
27	98.5	5.2	563	2 F64130	glucose-6-phosphat
28	98.5	5.2	2409	1 A60979	vesican precursor
29	98	5.2	3844	2 T18402	asparagine/asparta

30	97.5	5.1	1275	2 B28096	line-1 protein ORF
31	97	5.1	270	2 H83327	exodeoxyribonuclea
32	96.5	5.1	658	2 T33568	hypothetical prote
33	96.5	5.1	996	2 S46812	hypothetical prote
34	96.5	5.1	3191	2 T22945	hypothetical prote
35	96	5.0	289	2 AG3260	exodeoxyribonuclea
36	95	5.0	267	2 C97690	exodeoxyribonuclea
37	95	5.0	267	2 AH2915	exodeoxyribonuclea
38	95	5.0	1280	2 B34087	hypothetical prote
39	94.5	5.0	547	2 T30704	rifampicin resista
40	94	4.9	255	2 E70166	exodeoxyribonuclea
41	93.5	4.9	295	2 S34544	hypothetical prote
42	93.5	4.9	388	2 T47530	hypothetical prote
43	93.5	4.9	892	2 T50985	related to transcr
44	93	4.9	549	1 NDEC	glucose-6-phosphat
45	93	4.9	549	2 H91254	glucosephosphate 1

ALIGNMENTS

RESULT 1

T27272
hypothetical protein Y6303A.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27272

R:White, S.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20335

A:Accession: T27272

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-362 <MILL>

A:Cross-references: EMBL:AL032652; PIDN:CAA21707.1; GSPDB:GN00019; CESP:Y63D3A.4

A:Experimental source: clone Y63D3A

C:Genetics:

A:Gene: CESP:Y63D3A.4

A:Map position: 1

A:introns: 18/1; 43/1; 116/3; 148/2; 225/3; 320/2

Query Match 22.1%; Score 420; DB 2; Length 362;
Best local Similarity 29.3%; Pred. No. 2.4e-26;

Matches 108; Conservative 75; Mismatches 142; Indels 44; Gaps 10;

QY	7	LEGGRAAEDE-----GPEYKRRLLCV-----EFASVACDAVAOCFLAEN 50
DB	11	IEAKRKMSQEDSEVEETILDEPEQGLNSSSDQKLHEPAIITATDEAFQSLIDV 70
QY	51	DWMEBALNSYEPPEVESALERRPETISEPKTYVDLTNEETDSTSKISPEEDTOEN 110
DB	71	DWLKALDVFY-----GSEAFAPARSAAVAGASSMASGAAMVTAEDLK--- 116
QY	111	GSMFSLITWNIDGLDNLNLSERAGVCSYLALSPVITLQEVIPPIYSLKKRSSNEI 170
DB	117	GFEVSYMSWNIDGLDGRSLTRKRAVAHIVKNVPDILFLQEVYGRDLAPIDKLSVKI 176
QY	171	ITGHEE-GYFTALMLKKSRYKLSOELIPPSRKMRNLCLVHNVSGNELCLMTSHLS 229
DB	177	YYSNKGQYITATLVSR-MFDVKKHDVIFHONSGMTRTQLIEGSLGGLKFLVLTHTLS 235
QY	230	TRCHAAERNQMLKMLKKMQE--APESATVIFAGDTNLDRREVTFCGLPNNIVDWEF 286
DB	236	TREHRQRCKQGFQCMKDVREIIAQNPGALVFFGGDLNRDEVSR---VPDGVADAWA 292
QY	287	LGRPKHCQYTWDTQMSNLGCTACKLREDRIFFRAAAEAGHTIPSLDLGLGKL-DCG 345
DB	293	AGSDNKKKFTFDTFKNDKNGFGAKMRDRLYW-----SGPLDKVKTLEGRQRISCL 347
QY	346	RPSDHWGL 354
DB	348	CFPSDHWAI 356

```

RESULT 3
T40792
hypothetical protein SPAC9B6.11c - fission yeast (Schizosaccharomyces pombe)
C:Species: Schizosaccharomyces pombe
C:Date: 03-Dec-1999 #sequence_rev1500 03-Dec-1999 #text_change 09-Jun-2000
C:Accession: T40792
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volkart, G.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z21875
A:Accession: T40792
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-502 <MOO>
A:Cross-references: EMBL:AL049769; PTDN:CAB42372.1; GSPDB:GN00067; SPDB:SPAC9B6.11c
A:Experimental source: strain 972h-; cosmid c9B6
C:Genetics:
A:Gene: SPDB:SPAC9B6.11c
A:Map position: 2
A:Introns: 76/1
C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC9B6.11c

Query Match      5.9%; Score 113; DB 2; Length 502;

```

QY	51	DWMEERALNSFEED-----EVEESALRRREPTISEPTVYDLENEETJDSBTSK--	99
Dp	22	DYRLSKLAIISFDQKKNNKMAIDIPVKKGSAKGPPSFVTPR-YIEKQOKKLEKMAKAA	80
QY	100	---ISPSEDTQOEUGSM-----SLTWNIDGIDLNNL-----	129
Dp	81	RKPIAPSPNADPENPTNFIKREMLSPYVAPFETEKSAIDITIMTYN--LAQTNIRSM	138
QY	130	-----SEARAGVSYLALSPDYLIFQEV-----IPPYYSYLKKRSSNYEITGHE	175
Dp	139	FPHSGEALKKNNRSMLANELTYTSPILGICQOEVDAEVPRFYKKL-LGGELYELHFIKG	197
QY	176	EGYTATAML-KKSRYKLKSOEIIFF-----PSTKMMRILCY----HYVSSNNEICJM	223
Dp	198	EKGTHGIMEIFEKKSSIFKKVVDLTYYDDHDELPRMNTKNTGGCVRLERVDPPSRGFELA	257
QY	224	TSLSHETSGHAERNNOLKYNLK--KMOEAPESATVIFAGDTN-----LRDREV	270
Dp	258	TTHLFWHYGSEYERLROGAILLYKEVNNKMAOSHPRMVFYIADGFTNTERPDTNFRALTYRPL	317
QY	271	TRCGGLPNNIYDVMEFLGPKRHCQYTW-----DTOMNSNLGTATAC-----	311
Dp	318	SIC-----ORANDILE--RSMNYVFGESSELEEKMASTKTENDSNEDEKCECCSSSTSS	368
QY	312	-----KLFREDRI	318
Dp	369	VPESTASTPRKRIHLVQNDYVPHYRSPFYQOHEQNVPFLSYSGYKLVHPENAKNTDHP	428
QY	319	FFR-AAAEEGHI-----IPRSIDLGLLEKLDGCR	346
Dp	429	AFTWMAHYQCHLDYIEFWMNRDTSIQEPRENOVEGIKLLALRYVLPSEMKAEAPLE-GR	487
QY	347	FPSDHWGLCNLDII 361	
Dp	488	YPSDHVALMANVOIY 502	

RESULT 4
D95853
conserved hypothetical protein [imported] - *Sinorhizobium meliloti* (strain 1021) mega
C:Species: *Sinorhizobium meliloti*
C:Date: 24-Aug-2001 #sequence-revision 24-Aug-2001 #text-change 30-Sep-2001
C:Accession: D95853
R:Finan, T.M.; Weldner, S.; Wong, K.; Buhrmester, J.; Chaturvedi, P.; Vorholter, F.J.; Herr
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb *PSyMB* megaplasmid from the N2-fixing e
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95853
A:Molecule type: DNA
A:Residues: 1-252 <KOR>
A:Cross-references: GB:AL591985; PID:CAC48492.1; PID:g15139964; GSFDB:GN00167
A:Experimental source: Strain 1021, megaplasmid *PSyMB*
R:Galbreath, F.; Finan, T.M.; Long, S.R.; Punter, A.; Abola, P.; Ampe, F.; Barloy-Hubbl
pela, D.; Chaturvedi, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebaull, P.; Vandenberg, M.; Vorholter, F.J.; Weldner, S.; Wells, D.H.; Wong, K.; Yen,
A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SmbD20092
A:Genome: plasmid

Query Match	5.9%;	Score 111.5;	DB 2;	Length 252;
Best Local Similarity	24.4%;	Pred. No. 0.15;		
Matches	61;	Conservative	36;	Mismatches 96;
				Indels 57;
				Gaps 12;

QY 279 -----NIDYDWPEEL-GKP-----KHCQ-----YTWIDQMSNLGITR 309
 |||||
 Db 242 KETIQEAGLVDSWOLHGKODLAVIARSLPQOLLKCTTCDLSLNTWRAROPD-----E 297
 |||||
 QY 310 ACTLRF---DRIFRAAECHIIIPRSLDILGLEKLDGCGFPSPDHWGLCNLDIT 361
 |||||
 Db 298 ACRIDYALIPEDFLQTVDAVGRTER-----IPHLDCS--VSDHEFYSCTLLIIV 344
 |||||

RESULT 8

hypothetical protein ECs0868 [imported] - *Escherichia coli* (strain O157:H7, substrain R
 C:Species: *Escherichia coli*
 C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
 C:Accession: D90737
 R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.;
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
 Nucleic Acids Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
 A:Reference number: A96929; MUID:21156231; PMID:11258796
 A:Accession: D90737
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-253 <HAY>
 A:Cross-references: GB:BA000007; PTDN:BA34291.1; PTD:q13360327; GSPDB:GN00154
 A:Experimental source: strain O157:H7, substrain RMD 0509952
 C:Genetics:
 C:Gene: ECs0868

RESULT 9

hypothetical protein ybhp [imported] - *Escherichia coli* (strain O157:H7, substrain EDL95
C:Species: *Escherichia coli*
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 14-Sep-2001
C:Accession: E85587
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dialananta, E.; Potamousis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:1120651
A:Accession: E85587
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 17253 <STO>
A:Cross-references: GB:AE005174; NID:g12513779; PIDN:AAG5161.1; GSPDB:GN00145; UWGP:2107
A:Experimental source: strain O157:H7, substrain EDL953
C:Genetics:
C:Gene: ybhp

RESULT 10

[illegible]

RESULT 11

NE3001
T46340

AD115
 Hypothetical protein Ime0323 [Imported] - *Listeria monocytogenes* (strain EGD-e)
 C:Species: *Listeria monocytogenes*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AD115
 R:Glaser, P.; Fraigneul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurget, O.; Entlian, K.D.; Fsihl, H.; Jones, L.M.; Karst, U.
 Science 294, 849-852, 2001
 A:Authors: Kretz, J.; Kuhn, M.; Kunst, F.; Kurupkat, G.; Madueno, E.; Maitournam, A.; Mak, C.; Schlunegger, T.; Simoes, N.; Tlarriz, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AD115
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-256 <GLU>
 A:Cross-references: GB:IM003210; PIDN:CAD00850.1; PID:g16409687; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetics:
 A:Gene: Ime0323

Query Match	5.4%	Score 103.5	DB 2	length 256
Best Local Similarity	25.4%	Pred. No. 0.7		
Matches	45	Conservative	33	Mismatches 66; Indels 33; Gaps 9
QY	113	MESLTITWNIDGADLNNLSERARG-----VCSYLTATSPDYVFLDEVIPIPYSLTKRS	165	
	: :	:	: :	:
DB	1	MSVYTFNT---RFDDTSRRKSKSWELRKTLLTSLDKYGMDEMGVEPILLQPMRDKKMG	57	
QY	166	S-NYEII--TGHEEGYFSAIMLKSRVKL-----KSQEIIPBSTKMMRNLLCY	211	
	: : : :	: : :	: : : :	: : :
DB	58	DMDYGVGRDGEKEGFAVFNSTRFLLOEGHFWLSETPDVPEIHSHTAMP-R-ICWA	116	
QY	212	--HVNVSGNELMLNLSHLESTGHAEE--RMNDLKKVLLKKMGAPASAVYIFGDN	264	
	: : : : :	: : : :	: : :	: : :
DB	117	GRFASDGGQFIYFNTHD---HISEEARLRASQLLKKAAITANSPIVLIIGDN	169	

RESULT 15
 T24066
 exonuclease III homolog - *Caenorhabditis elegans*
 C:Species: *Caenorhabditis elegans*
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
 C:Accession: T24066; T42391
 R:Kershaw, J.
 submitted to the EMBL Data Library, October 1996
 A:Reference number: Z19837
 A:Accession: T24066
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: DNA
 A:Residues: 1-288 <MW>
 A:Cross-references: EMBL:Z81108; PIDN:CAB03235.1; GSPDB:GN00019; CESP:R0983.1
 A:Experimental source: clone R09B3
 R:Tremblay, S.; Masson, J.Y.; Ramotar, D.
 submitted to the EMBL Data Library, November 1997
 A:Description: The exonuclease III family of DNA repair enzymes has a new homolog in *Ce*
 A:Reference number: Z22150
 A:Accession: T42391
 A:Status: preliminary; translated from GB/EMBL/DBD
 A:Molecule type: mRNA
 A:Residues: 1-288 <TR>
 A:Cross-references: EMBL:AF034258; PIDN:AAC82328.1
 C:Genetics:
 A:Gene: R0983.1
 A:Map position: 1
 A:introns: 83/3; 192/3; 264/3
 C:Superfamily: exodeoxyribonuclease III

```

OY 72 ERREPTESEKTYVDLNNEEETDSTSTKISISEPTQOENGSMFLIMNIDGDLNNLSE 131
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 3 KKAEEADAPK-----LASIFTKKVKRAE--EDNNOKSMKFVCNNAAGL----- 44

OY 132 RA--RGVCSYALYSPDVFILQEV---IPPYYSYLKRRSSNEYEITGHEGYFTYIML 184
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 45 RACVYKSDPFREVLAEEDPLVFLETCKKEMPEMEETFKNTKTLVSTENGGSYACVGL 104
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 185 KKSRYVTLKSOELIPPPSTKMMRNLLC-----YHVNNSGNEJLCLMSTHLESTRGHA 234
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 105 LSKCAPMKVHVHGIGDPEDTAGRLIIAEFSKFPYITGAVPNSGAKLV---NLE----- 154
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

OY 235 AERNQMLKAVLKKMOEAPESATVIFAGDTNLRDEEV 270
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 155 -KRGWEKULTTEKKEMDEKKPVYIIGGLNVAHNHEI 189
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: November 14, 2002, 10:28:15
Job time : 16.5673 secs

Query Match	5.4%;	Score 103;	DB 2;	Length 288;
Best*Local Similarity	21.8%;	Pred. No. 0.91;		
Matches	47;	Conservative	33;	Mismatches 90;
				Indels 46;
				Gaps 8;

[illegible]

Search completed: November 14, 2002, 10:28:15
Job time : 16.5673 secs

Query Match	5.4%;	Score 103;	DB 2;	Length 288;
Best*Local Similarity	21.8%;	Pred. NO. 0.91;		
Matches	47;	Conservative	33;	Mismatches 90;
				Indels 46;
				Gaps 8;

QY	241	LKAVLKMKQEPASATVYTFEGDITNLDRREYTRCGGLPNNIIVDWEELGKRKHCQYTWDTQ	300
QY	241	LKAVLKMKQEPASATVYTFEGDITNLDRREYTRCGGLPNNIIVDWEELGKRKHCQYTWDTQ	300
QY	301	MNSNLGITTACCKLRFDPIFPRAAAEEGHITPRSIDLIGLEKIDCGRPSPDHWGILCNLDI	360
Db	301	MNSNLGITTACCKLRFDPIFPRAAAEEGHITPRSIDLIGLEKIDCGRPSPDHWGILCNLDI	360
QY	361	IL 362	
QY	361	IL 362	
Db	361	IL 362	
+RESULT 3			
XX	AA041627		
XX	AA041627	standard; Protein; 372 AA.	
AC	AA041627;		
XX			
DT	22-OCT-2001	(first entry)	
XX			
DE	Human polypeptide SEQ ID NO 6558.		
XX			
KW	Human; neurotropic; immunosuppressant; cytosstatic; gene therapy; cancer;		
KW	peripheral nervous system; neuropathy; central nervous system; CNS;		
KW	Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;		
KW	amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;		
KW	chemokine; thrombolytic; drug screening; arthritis; inflammation;		
KW	leukemia.		
XX			
OS	Homo sapiens.		
XX			
PN	WO200153312-A1.		
XX			
PD	26-JUL-2001.		
XX			
PF	26-DEC-2000; 2000MO-US34263.		
XX			
PR	21-JAN-2000; 2000US-0488725.		
PR	25-APR-2000; 2000US-0552317.		
PR	09-JUL-2000; 2000US-0598042.		
PR	19-JUL-2000; 2000US-0620312.		
PR	03-AUG-2000; 2000US-0653450.		
PR	14-SEP-2000; 2000US-0662191.		
PR	19-OCT-2000; 2000US-0693036.		
XX	29-NOV-2000; 2000US-0727344.		
XX			
PA	(HYSE-) HYSEQ INC.		
XX			
PI	Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;		
PI	Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;		
PI	Zhao QA, Zhou P, Goodrich R, Drmanac RT;		
DR	WPI; 2001-442253/47.		
XX	N-PSDB; AAI60783.		
PT	Novel nucleic acids and polypeptides, useful for treating disorders		
PT	such as central nervous system injuries -		
XX			
PS	Example 2; SEQ ID NO 6558; 10078pp; English.		
XX			
CC	The invention relates to human nucleic acids (AA157798-AA161369) and		
CC	the encoded polypeptides (AA038642-AA042213) with neurotropic,		
CC	immunosuppressant and cytosstatic activity. The polynucleotides are useful		
CC	in gene therapy. A composition containing a polypeptide or polynucleotide		
CC	of the invention may be used to treat diseases of the peripheral nervous		
CC	system, such as peripheral nervous injuries, peripheral neuropathy and		
CC	localised neuropathies and central nervous system diseases, such as		
CC	Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic		
CC	lateral sclerosis, and Shy-Drager Syndrome. Other uses include the		
CC	utilisation of the activities such as: Immune system suppression,		
CC	Activin/inhibin activity, chemotactic/chemokine activity, haemostatic		
CC	and thrombolytic activity, cancer diagnosis and therapy, drug screening,		

[illegible]

CC This sequence represents the topoisomerase II binding protein (TopBP) of
 CC the invention. The TopBP protein is useful as an anticancer agent. TopBP
 CC can be used as the target molecule for anticancer agent.

XX Sequence 392 AA.

Query Match 100.0%; Score 1902; DB 20; Length 392;
 Best Local Similarity 100.0%; Pred. No. 3,7e-188;
 Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSLGEGREAAEEGEPEVKRRLLCVERASVASCDAVAACFLAENDMEMERALNS 60
 DB 31 MELGSLGEGREAAEEGEPEVKRRLLCVERASVASCDAVAACFLAENDMEMERALNS 90
 QY 61 YFPPVEESALERRPETISEPKTYVDLTNEETDSTSKISPEBDTQENGSMFSLITWN 120
 DB 91 YFPPVEESALERRPETISEPKTYVDLTNEETDSTSKISPEBDTQENGSMFSLITWN 150
 QY 121 IDGLDNLNLSERAGVCSYALYSPDYVIFLOEVIPPYSYLKKRSSNVEITGHEGYFT 180
 DB 151 IDGLDNLNLSERAGVCSYALYSPDYVIFLOEVIPPYSYLKKRSSNVEITGHEGYFT 210
 QY 181 AIMLKSRVRLKSOEIIIPPSTKMMRNLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 DB 211 AIMLKSRVRLKSOEIIIPPSTKMMRNLCVHNVSGNELCLMTSHLESTRGHAARMNQ 270
 QY 241 LKWLKMOEAPESATVIFAGDTNLRDREVTGCGLPNNIVDWEFLGKPKHCQYTWDTQ 300
 DB 271 LKWLKMOEAPESATVIFAGDTNLRDREVTGCGLPNNIVDWEFLGKPKHCQYTWDTQ 330
 QY 301 MNSNLGITTAACKLRFDRIFFRAAAEHGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 360
 DB 331 MNSNLGITTAACKLRFDRIFFRAAAEHGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 390

QY 361 IL 362
 DB 391 IL 392

RESULT 5

AAAB53403 ID AAB53403 standard; Protein; 369 AA.

AAAB53403;

09-MAR-2001 (first entry)

Human colon cancer antigen protein sequence SEQ ID NO:943.

Human; colon cancer; colon cancer antigen; diagnosis; detection;
 identification; cytostatic; cardioactive; neuroprotective; vulnerary;
 immunomodulatory; muscular; gynaecological; gastrointestinal;
 neoplastic; antineoplastic; antibacterial; gene therapy; wound;
 neutral disorder; immune system disorder; muscular disorder;
 reproductive disorder; gastrointestinal disorder; renal disorder;
 infectious disease; cardiovascular disorder.

Homo sapiens.

WO200055351-A1.

21-SEP-2000.

08-MAR-2000; 2000WO-US05883.

12-MAR-1999; 99US-0124270.

(HUMA-) HUMAN GENOME SCI INC.

Rosen CA, Ruben SM;

WPI, 2000-587534/5.

N-PSDB; AAC98160.

XX Colon cancer associated gene sequences, referred to as colon cancer
 PT antigens, useful for the treatment, prevention, and diagnosis of colon
 PT disorders such as colon cancer -

Claim 11: Page 1508-1509; 2104pp; English.

AAAC9791 to AAC98763 encode the human colon cancer associated proteins,
 called human colon cancer antigens, given in AAB53224 to AAB54006. The
 human colon cancer antigens can have cytostatic, cardioactive, muscular,
 neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
 vulnerary, nephrotropic, antineoplastic and antibacterial activities, and
 can be used in gene therapy. The colon cancer antigen polynucleotides,
 proteins and antibodies to the proteins are useful for the prevention,
 treatment and diagnosis of colon disorders, such as colon cancer. The
 polynucleotides may be used in diagnostics and research, such as for
 chromosome identification, and as hybridisation probes. The proteins
 may also be used to prevent diseases such as neutral disorders, immune
 CC system disorders, muscular disorders, reproductive disorders,
 CC gastrointestinal disorders, wounds, renal disorders, infectious
 CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
 CC AAB54007 represent sequences used in the exemplification of the present
 CC invention.

XX Sequence 369 AA;

Query Match 99.6%; Score 1894; DB 21; Length 369;
 Best Local Similarity 99.7%; Pred. No. 2.2e-187;
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSLGEGREAAEEGEPEVKRRLLCVERASVASCDAVAACFLAENDMEMERALNS 60
 DB 8 MELGSLGEGREAAEEGEPEVKRRLLCVERASVASCDAVAACFLAENDMEMERALNS 67
 QY 61 YFPPVEESALERRPETISEPKTYVDLTNEETDSTSKISPEBDTQENGSMFSLITWN 120
 DB 68 YFPPVEESALERRPETISEPKTYVDLTNEETDSTSKISPEBDTQENGSMFSLITWN 127
 QY 121 IDGLDNLNLSERAGVCSYALYSPDYVIFLOEVIPPYSYLKKRSSNVEITGHEGYFT 180
 DB 128 IDGLDNLNLSERAGVCSYALYSPDYVIFLOEVIPPYSYLKKRSSNVEITGHEGYFT 187
 QY 181 AIMLKSRVRLKSOEIIIPPSTKMMRNLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 DB 188 AIMLKSRVRLKSOEIIIPPSTKMMRNLCVHNVSGNELCLMTSHLESTRGHAARMNQ 247
 QY 241 LKWLKMOEAPESATVIFAGDTNLRDREVTGCGLPNNIVDWEFLGKPKHCQYTWDTQ 300
 DB 248 LKWLKMOEAPESATVIFAGDTNLRDREVTGCGLPNNIVDWEFLGKPKHCQYTWDTQ 307
 QY 301 MNSNLGITTAACKLRFDRIFFRAAAEHGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 360
 DB 308 MNSNLGITTAACKLRFDRIFFRAAAEHGHIIPRSLDLGLEKLDCCGRFPSDHWGLLCNLDI 367
 QY 361 IL 362
 DB 368 IL 369

RESULT 6

AAAB93674 ID AAB93674 standard; Protein; 362 AA.

AAAB93674;

26-JUN-2001 (first entry)

Human protein sequence SEQ ID NO:13210.

Human; primer; detection; diagnosis; antisense therapy; gene therapy.

Homo sapiens.

PN EPI074617-A2.
 XX
 PD 07-FEB-2001.
 XX
 PF 28-JUL-2000; 2000EP-0116126.
 XX
 PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 DR WPI: 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs -
 PS
 PS Claim 8; SEQ ID 13210; 2537bp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the combination of the oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH13628 and AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632 represent oligonucleotides, all of which are used in the exemplification of the present invention.

XX Sequence 362 AA:

Query Match 99.5%; Score 1893; DB 22; Length 362;
 Best Local Similarity 99.7%; Pred. No. 2,8e-187;
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSCLEGGREAAEEGPEVKKRRLLCYEFASVASCDAVAQCLAEIENDWEMERALNS 60
 DB 1 MELGSCLEGGREAAEEGPEVKKRRLLCYEFASVASCDAVAQCLAEIENDWEMERALNS 60
 QY 61 YFEPPEVESALERRPETISEPKTYVDLTNEETDSTSKISPSDQOENGSMFSLITWN 120
 DB 61 YFEPPEVESALERRPETISEPKTYVDLTNEETDSTSKISPSDQOENGSMFSLITWN 120
 QY 121 IDGIDLNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEEGYFT 180
 DB 121 IDGIDLNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEEGYFT 180
 QY 181 AIMKKRSVKLSKSOEIIIPFPSTKMMRNLLCYHVVNSGNELCMLTSHLESTRGHAAREMNQ 240
 DB 181 AIMKKRSVKLSKSOEIIIPFPSTKMMRNLLCYHVVNSGNELCMLTSHLESTRGHAAREMNQ 240
 QY 241 LKAVLKKMOEAPESATYIFAGDTNLRDREVT RCGGLPNNIYDVWEFLGKPKHCQYTDQ 300
 DB 241 LKAVLKKMOEAPESATYIFAGDTNLRDREVT RCGGLPNNIYDVWEFLGKPKHCQYTDQ 300

DB 241 LKAVLKKMOEAPESATYIFAGDTNLRDREVT RCGGLPNNIYDVWEFLGKPKHCQYTDQ 300
 QY 301 MNSNIGITAAKCLRFDRIFPRAAAEEGHIIPRSIDLGLGKLDGGRPSDHWGLCNLDI 360
 DB 301 MNSNIGITAAKCLRFDRIFPRAAAEEGHIIPRSIDLGLGKLDGGRPSDHWGLCNLDI 360
 QY 361 IL 362
 DB 361 IL 362

RESULT 7

AAV03181
 ID AAV03181 standard; protein; 306 AA.

AC AAV03181;

DT 16-JUN-1999 (first entry)

DE Topoisomerase II binding protein.

KW Topoisomerase II binding protein; TopBP; anticancer agent.

OS Homo sapiens.

PN JP11075856-A.

PD 23-MAR-1999.

PE 17-SEP-1997; 97JP-0251544.

PR 17-SEP-1997; 97JP-0251544.

PA (CHUGAI) CHUGAI PHARM CO LTD.

PA (TSURU/) TSURUO T.

DR WPI: 1999-257704/22.

PT New Topoisomerase II- binding protein - useful as an anticancer agent

PS Claim 1; Page 12-13; 28pp; Japanese.

CC This sequence represents the topoisomerase II binding protein (TopBP) of the invention. The TopBP protein is useful as an anticancer agent. TopBP can be used as the target molecule for anticancer agent.

XX Sequence 306 AA:

Query Match 84.6%; Score 1609; DB 20; Length 306;
 Best Local Similarity 100.0%; Pred. No. 6e-158;
 Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 57 AINSYFEPVESALERRPETISEPKTYVDLTNEETDSTSKISPSDQOENGSMFSL 116
 DB 1 AINSYFEPVESALERRPETISEPKTYVDLTNEETDSTSKISPSDQOENGSMFSL 116
 QY 117 ITWNIDGLDNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEE 176
 DB 117 ITWNIDGLDNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEE 176
 QY 61 ITWNIDGLDNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEE 120
 DB 61 ITWNIDGLDNNLSERARVC SYALYSPDVI FLOEVIPYYSYLKKRSSNVEIITGHEE 120
 QY 177 GYFTALMLKRSVKLSKSOEIIIPFPSTKMMRNLLCYHVVNSGNELCMLTSHLESTRGHA 236
 DB 177 GYFTALMLKRSVKLSKSOEIIIPFPSTKMMRNLLCYHVVNSGNELCMLTSHLESTRGHA 236
 QY 237 RANOLKMKVKKMOEAPESATYIFAGDTNLRDREVT RCGGLPNNIYDVWEFLGKPKHCQY 296
 DB 237 RANOLKMKVKKMOEAPESATYIFAGDTNLRDREVT RCGGLPNNIYDVWEFLGKPKHCQY 296
 QY 297 WDTOMNSNIGITAAKCLRFDRIFPRAAAEEGHIITRSIDLGLGKLDGGRPSDHWGLLC 356
 DB 297 WDTOMNSNIGITAAKCLRFDRIFPRAAAEEGHIITRSIDLGLGKLDGGRPSDHWGLLC 356
 QY 356 WDTOMNSNIGITAAKCLRFDRIFPRAAAEEGHIITRSIDLGLGKLDGGRPSDHWGLLC 356
 DB 356 WDTOMNSNIGITAAKCLRFDRIFPRAAAEEGHIITRSIDLGLGKLDGGRPSDHWGLLC 356

QY 357 NLDIIL 362
 DB 301 NLDIIL 306
 Y RESULT 8
 ID AAY56020 standard; Protein; 370 AA.
 AC AAY56020;
 DT 15-MAR-2000 (first entry)
 DE Mouse CD40 receptor associated protein.
 KW Antiartherosclerotic; antiarthritic; neuroprotective; dermatological;
 KW immunosuppressive; antiinflammatory; immunosuppressive; antiallergic;
 KW human; CD40 receptor associated protein; CRAP; cytoplasmic domain;
 KW tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;
 KW TNF receptor associated factor; TRAF; modulator; signalling pathway;
 KW diagnosis; NF-kappaB; Jun; kinase; atherosclerosis; multiple sclerosis;
 KW arthritis; systemic lupus erythematosus; graft rejection; allergy;
 KW graft versus host disease; autoimmune disease.
 XX
 OS Mus musculus.
 XX
 PN WO955859-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 28-APR-1999; 99WO-EP03025.
 XX
 PR 29-APR-1998; 98EP-0201392.
 XX
 PA (VLA-) VLAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
 XX
 PI Pype SMC, Remacle JEFUG, Huylebnoeck DFE;
 XX
 DR WPI: 2000-062029/05.
 DR N-PSDB; AAZ47119.
 XX
 PT Novel proteins used to treat inflammatory diseases, NF-kappaB related
 PT diseases and for improvement of anti-tumor treatments
 PS Claim 3; Page 41-43; 48pp; English.
 CC This sequence represents the mouse CD40 receptor associated protein
 CC (CRAP). CRAP is a functional protein capable of interacting with the
 CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
 CC factor (TNF) receptor superfamily such as CD30 and TNF receptor 1, where
 CC the protein has no homology to TNF receptor associated factor (TRAF)
 CC proteins. The CD40 binding proteins can be used as modulators of the
 CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
 CC CD40-related, NF-kappaB related and/or Jun (kinase)-related diseases,
 CC and for the improvement of anti-tumour diseases. Diseases which may be
 CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
 CC lupus erythematosus, graft rejection, graft versus host disease, systemic
 CC and autoimmune disease. The proteins can be used to sensitize tumour
 CC cells to anti-tumour treatments and to screen for compounds which
 CC interfere with the interaction of the proteins with other protein
 CC components of the TRAF, CD40 or NF-kappaB related pathway.
 CC
 XX
 SQ Sequence 370 AA;
 Query Match 66.1%; Score 1258; DB 21; Length 370;
 Best Local Similarity 68.4%; Pred. No. 2,1e-121;
 Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;
 QY 12 EAAREEPEVKRRRLLCVERASASDAVAQCFLAENDWEMERALNSYFEPVESAL 71
 DB 24 EAQAE-EDRKRRRLQCLGFALVGGCDPTMPVSVLENEMOTOKALSAVFLPNDQGW 82
 QY 72 ERRETSISEPKTYVDLTNETTDTSTSKISPSEDTQOENGSMSLITWNIDGLNNLS 131

DB 83 PROPTSFKSEAYVDLTNETDANDTIIIEASPS-GTPLEDSSTISFTWNIDGLDGNLPE 141
 QY 132 RARGVCSYALYSPDVIFLQEVIPPPYSSYLKRRSSNYEITLHGEGYFTAILMKRSYKL 191
 DB 142 RARGVCSCLALYSPDVVFLQEVIPPCAYLKRRASYYITIGNEGFTAILLKGGYKF 201
 QY 192 KSQELTTPPSTKMMRNLLCVAVNWSGNELCLMTSHLESTRGHAERNOLKMYLKKMOEA 251
 DB 202 KSQELTTPPNTKMMRNLLCVAVNSLGNELCLMTSHLESTRGHAERNOLKMYLKKMOEA 261
 QY 252 PESATVIFAGDNTLRDERVTRCGGLPNNIYDVWFELGPKPHCOYTWTQNNNSNGITAA 311
 DB 262 PDSTTVIFAGDNTLRDEVIKCGGLPDNVDPAMFELCKPHCOYTWTQKANNLRIPAA 321
 QY 312 KLKRDRIFFRAAEHIIIPRSIDLGLERKIDGGRPSDHMLLCNDIIL 362
 DB 322 KHRDRIFFR--AREGHILPQSLDLVGLERKIDGGRPSDHMLLCITLVNL 370
 Y RESULT 9
 ID ABG22067 standard; Protein; 263 AA.
 AC ABG22067;
 DT 18-FEB-2002 (first entry)
 DE Novel human diagnostic protein #22058.
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 XX
 OS Homo sapiens.
 XX
 PN WO200175067-A2.
 XX
 PD 11-OCT-2001.
 XX
 PF 30-MAR-2001; 2001WO-US08631.
 XX
 PR 31-MAR-2000; 2000US-0540217.
 PR 23-AUG-2000; 2000US-0649167.
 XX
 PA (HYSE-) HYSBO INC.
 XX
 PI Drmanac RT, Liu C, Tang YT;
 XX
 DR WPI: 2001-639362/73.
 DR N-PSDB; AAS86254.
 XX
 PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT biologically -
 PS Claim 20; SEQ ID No 52426; 103pp; English.
 CC The invention relates to isolated polynucleotide (I) and
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
 CC and gene mapping, and in recombinant production of (II). The
 CC polynucleotides are also used in diagnostics as expressed sequence tags
 CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity

CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 263 AA;

Query Match 63.0%; Score 1199; DB 22; Length 263;
 Best Local Similarity 98.7%; Pred. No. 1.6e-115;
 Matches 226; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 133 AGGVCSTLALSPDYFLQEVIPYYSYLRKRSNVEITGHEEGYFTAIMLKSRVLT 192
 CC |||||||
 CC 1 AAGVCSTLALSPDYFLQEVIPYYSYLRKRSNVEITGHEEGYFTAIMLKSRVLT 60

OY 193 SOEIIPEPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMMOLKMYLKKMOEAP 252
 CC |||||||
 CC 61 SOEIIPEPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMMOLKMYLKKMOEAP 120

OY 253 ESATVFAGDTNLRDEYTRCGGLPNNTVDVWEFLGPKHCQYTWDTOMNSNLGITACK 312
 CC |||||||
 CC 121 ESATVFAGDTNLRDEYTRCGGLPNNTVDVWEFLGPKHCQYTWDTOMNSNLGITACK 180

OY 313 LRDRIFFRAAAEHGHIIIPRSIDLGLLEKLDGRRPSDHMGLLCNDII 361
 CC |||||||
 CC 181 LRDRIFFRAAAEHGHIIIPRSIDLGLLEKLDGRRPSDHMGLLCNDII 229

Db 181 LRDRIFFRAAAEHGHIIIPRSIDLGLLEKLDGRRPSDHMGLLCNDII 229

RESULT 10

ABG22068
 ID ABG22068 standard; Protein; 311 AA.

AC ABG22068;

DT 18-FEB-2002 (first entry)

DE Novel human diagnostic protein #22059.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;

KM food supplement; medical imaging; diagnostic; genetic disorder.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HXSF-) HXSQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR N-PSDB; AAS86255.

PT New isolated polynucleotide and encoded polypeptides, useful in
 PT diagnostics, forensics, gene mapping, identification of mutations
 PT responsible for genetic disorders or other traits and to assess
 PT biodiversity

XX Claim 20; SEQ ID No 52427; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and

CC polypeptide (II) sequences. (I) is useful as hybridisation probes,

CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome

CC and gene mapping, and in recombinant production of (II). The

CC polynucleotides are also used in diagnostics as expressed sequence tags

CC for identifying expressed genes. (I) is useful in gene therapy techniques
 CC to restore normal activity of (II) or to treat disease states involving
 CC (II). (II) is useful for generating antibodies against it, detecting or
 CC quantitating a polypeptide in tissue, as molecular weight markers and as
 CC a food supplement. (II) and its binding partners are useful in medical
 CC imaging of sites expressing (II). (I) and (II) are useful for treating
 CC disorders involving aberrant protein expression or biological activity.
 CC The polypeptide and polynucleotide sequences have applications in
 CC diagnostics, forensics, gene mapping, identification of mutations
 CC responsible for genetic disorders or other traits to assess biodiversity
 CC and to produce other types of data and products dependent on DNA and
 CC amino acid sequences. ABG00010-ABG30377 represent novel human
 CC diagnostic amino acid sequences of the invention.
 CC Note: The sequence data for this patent did not appear in the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pct_sequences.

SO Sequence 311 AA;

Query Match 60.5%; Score 1150.5; DB 22; Length 311;
 Best Local Similarity 66.3%; Pred. No. 2.2e-110;
 Matches 242; Conservative 13; Mismatches 29; Indels 81; Gaps 6;

OY 1 MELSGLEGGREAAEEGEPYKRRLLCYEFASVASCDAVAQCFLAENDMEERALNS 60
 CC |||||||

Db 25 MELSGCLYGFREAAEBEGEPYKRRLLCYEFASVASCDAVAQCFLAENDMEERALNS 84

OY 61 YFEPPEESALERREPETISEPKTYVDLTNETTDTSTSKISPSEDTQENGSMFSLITWN 120
 CC |||||||

Db 85 YFEPPEESALERREPETISEPKTYVDLTNETTDTSTSKISPSEDTQENGSMFSLITWE 144

OY 121 -IDGIDL-NMSEARQVCSYLALSPDYFLQEVIPYYSYLRKRSNVEITGHEGY 178
 CC |||||

Db 145 YLMGIDLKQSVQRARGVCSYLALXVLSLFLNSNV-----SYV----- 182

OY 179 FTAILMKRSVKLSQETIIPPPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERM 238
 CC |||||

Db 183 -----YTERM 187

OY 239 NQLKMYLKKMOEAPESATVIFAGDTNLRDEYTRCGGLPNNTVDVWEFLGPKHCQYTW 297
 CC |||||||

Db 188 NQLKMYLKKMOEAPESATVIFAGDTNLRDEYTRCGGLPNNTVDVWEFLGPKHCQYTW 246

OY 298 DTQMSNLGITACKLRDRIFFRAAAEHGHIIIPRSIDLGLLEKLDGRRPSDHMGLLCN 357
 CC |||||||

Db 247 DTQMSNLGITACKLRDRIFFRAAAEHGHIIIPRSIDLGLLEKLDGRRPSDHMGLLCN 306

OY 358 IDIIL 362
 CC |||||

Db 307 IDIIL 311

RESULT 11

AAV11868
 ID AAV11868 standard; Protein; 161 AA.

AC AAV11868;

DT 18-JUN-1999 (first entry)

DE Human 5' EST secreted protein SEQ ID No: 468.

KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide; prostate;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokine; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.

OS Homo sapiens.

PN WO9906550-A2.

XX

PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-IB01232.
 XX
 PR 01-AUG-1997; 97US-0905144.
 XX
 PA (GENSET) GENSET.
 XX
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 XX
 DR WPI; 1999-153780/13.
 XX
 DR N-PSDB; AAX40590.
 XX
 PT New isolated prostate-derived nucleic acids - used to develop
 PT products which may have cytokine, immune regulatory, haematopoiesis
 PT regulating, anti-inflammatory or tumour inhibition activity
 XX
 PS Claim 34; Page 595; 675pp; English.

CC AAX40438 to AAX40715 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins expressed in prostate, and encode the proteins
 CC given in AAY11716 to AAY11993 respectively. The proteins given represent
 CC the signal peptide and an N-terminal fragment of a secreted protein. The
 CC nucleic acid sequences can be used for producing secreted human gene
 CC products. They can also be used to develop products for diagnosis and
 CC therapy. The proteins obtained may have cytokine activity, cell
 CC proliferation and differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptides can be used for
 CC directing extracellular secretion of a polypeptide or the insertion of a
 CC polypeptide into a membrane, or importing a polypeptide into a cell.
 XX
 SQ Sequence 161 AA;

Query Match 44.0%; Score 837; DB 20; Length 161;
 Best Local Similarity 100.0%; Pred. No. 2.4e-78;
 Matches 161; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGGREAAEEGEPVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 60
 DB 1 MELGSCLEGGREAAEEGEPVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 60
 QY 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTSTKISPSDIOENGSMFSLITWN 120
 DB 61 YFEPPEESALERRPETISEPKTYVDLTNEETDSTSTKISPSDIOENGSMFSLITWN 120
 QY 121 IDGIDLNNLSERAGVCSYIALNSPDYIFIQEVIPIPYSTL 161
 DB 121 IDGIDLNNLSERAGVCSYIALNSPDYIFIQEVIPIPYSTL 161

RESULT 12
 AAY12175
 ID AAY12175 standard; Protein; 76 AA.
 XX
 AC AAY12175;
 XX
 DT 18-JUN-1999 (first entry)
 XX
 DE Human 5' EST secreted protein SEQ ID NO: 488.
 XX
 KW Human; secreted protein; EST; expressed sequence tag; diagnosis;
 KW forensic; gene therapy; chromosome mapping; signal peptide;
 KW upstream regulatory sequence; cytokine activity; cell proliferation;
 KW differentiation; haematopoiesis regulation; tissue growth regulation;
 KW reproductive hormone regulation; chemotactic; chemokinetic; haemostatic;
 KW thrombolytic; anti-inflammatory; tumour inhibition.

XX
 OS Homo sapiens.
 XX
 FN W09906554-A2.
 XX
 PD 11-FEB-1999.
 XX
 PF 31-JUL-1998; 98WO-IB01238.
 XX
 PR 01-AUG-1997; 97US-0905134.
 XX
 PA (GENSET) GENSET.
 XX
 PI Duclert A, Dumas Milne Edwards J, Lacroix B;
 XX
 DR WPI; 1999-153784/13.
 XX
 DR N-PSDB; AAX41008.
 XX
 PT New nucleic acids encoding human secreted proteins - obtained from
 PT cDNA libraries prepared from kidney, fetal kidney, dystrophic
 PT muscle, muscle and heart tissue
 XX
 PS Claim 34; Page 567; 622pp; English.

CC AAX40826 to AAX41093 represent 5' expressed sequence tags (ESTs) for
 CC human secreted proteins, and encode the proteins given in AAY01602 and
 CC AAY11994 to AAY12260, respectively. The proteins given represent the
 CC signal peptide and an N-terminal fragment of a secreted protein. The
 CC nucleic acid sequences can be used for producing secreted human gene
 CC products. They can also be used to develop products for diagnosis and
 CC therapy. The proteins obtained may have cytokine activity, cell
 CC proliferation/differentiation activity, haematopoiesis regulating
 CC activity, tissue growth regulating activity, reproductive hormone
 CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
 CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
 CC activity, tumour inhibition activity or other activities. The products
 CC can be used in forensic, gene therapy and chromosome mapping procedures.
 CC The sequences can also be used for obtaining corresponding promoter
 CC sequences. The nucleic acids encoding the signal peptide can be used
 CC for directing extracellular secretion of a polypeptide or the insertion
 CC of a polypeptide into a membrane, or importing a polypeptide into
 CC a cell.
 XX
 SQ Sequence 76 AA;

Query Match 19.8%; Score 376; DB 20; Length 76;
 Best Local Similarity 96.0%; Pred. No. 4.6e-31;
 Matches 72; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELGSCLEGGREAAEEGEPVKRRLLCVFASVASCDAVAQCFLAENDWEMERALNS 60
 DB 1 MELGSCLEGGREAAEEGEPVKRRLLCXEFXSVAASCDAAVAQCFLAXNEMERALNS 60
 QY 61 YFEPPEESALERRP 75
 DB 61 YFEPPEESALERRP 75

RESULT 13
 AAG26231
 ID AAG26231 standard; Protein; 404 AA.
 XX
 AC AAG26231;
 XX
 DT 17-OCT-2000 (first entry)
 XX
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 30610.
 XX
 KW Protein identification; signal transduction pathway; metabolic pathway;
 KW hybridisation assay; genetic mapping; gene expression control; promoter;
 KW termination sequence.
 XX
 OS Arabidopsis thaliana.

XX EPI033405-A2.
PN
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 18-MAY-1999; 99US-0134570.
PR 19-MAY-1999; 99US-0134768.
PR 20-MAY-1999; 99US-0134941.
PR 21-MAY-1999; 99US-0135124.
PR 24-MAY-1999; 99US-0135553.
PR 25-MAY-1999; 99US-0135629.
PR 27-MAY-1999; 99US-0136021.
PR 28-MAY-1999; 99US-0136592.
PR 01-JUN-1999; 99US-0136782.
PR 03-JUN-1999; 99US-0137222.
PR 04-JUN-1999; 99US-0137528.
PR 07-JUN-1999; 99US-0137502.
PR 08-JUN-1999; 99US-0137724.
PR 10-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 14-JUN-1999; 99US-0138847.
PR 16-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 17-JUN-1999; 99US-0139453.
PR 18-JUN-1999; 99US-0139452.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0145086.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145222.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.

PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145219.
PR 26-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148341.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.

PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.1%; Score 231; DB 21; Length 404;
Best Local Similarity 30.2%; Pred. No. 7.9e-15;
Matches 84; Conservative 39; Mismatches 109; Indels 46; Gaps 11;

QY 113 MSLITWNIDGIDLNLSRRAGVCSYLALYSPDVILOEVIPPYYSYLKKR-----SS 166
DB 138 ILSYVWPFREDLELN--LKMRAIGHLIQLHSPHLICPEVPEPEYIDIPFRSNMKKAVSC 194
QY 167 NEEIITGHEGEYFTALMKRSVKLKSOEIIFFPSTKMKRNLCYVNVSGNE-LCLMTS 225
DB 195 SVSVDVAVSRGYICML--SKLSGVKSFSSKFSGNSIMRELSTIAVEVPGKRPVAFATS 251
QY 226 HLES-----TRGHAAERMMOLKMWLKKMOEAPESATVIFAGTNNLRDREYTRCGG-- 275
DB 252 HLESPPGPPKMDOMFSRREVEDAKAEATFILLR--PNANVIFGDMNCD---KLDFK 304
QY 276 -LPNNIVDWWEFLGKPKHCQYTWDTOMNSNLGTFACKLRFPRIFFR----- 321
DB 305 PLPDKWVDVWEVL-KPGDLCFTYIDIKANPMLSGNNAALOKRDRITLICRLDDYKLGSIEMWG 363
QY 322 AAEEGHIIPRSIDLIG-LEKLDGGRFPSDHWGLLONT 358

Dh 364 KEAIPGSLYVEKKKVRGDIKKLEPVLPSDFGLLYTL 401

RESULT 15

AGG26230

ID AAG26230 standard; Protein; 426 AA.

XX

AC AAG26230;

XX

DT 17-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 30609.

XX

KW Protein identification; signal transduction pathway; metabolic pathway;

KW hybridisation assay; genetic mapping; gene expression control; promoter;

XX termination sequence.

XX

OS Arabidopsis thaliana.

XX

PN EP1033405-A2.

XX

PD 06-SEP-2000.

XX

PF 25-FEB-2000; 200EP-0301439.

XX

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 28-APR-1999; 99US-0130891.

PR 30-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 07-MAY-1999; 99US-0132487.

PR 11-MAY-1999; 99US-0132863.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137528.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139899.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.

PR 30-JUN-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 21-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 21-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148345.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

PR 17-AUG-1999; 99US-0149175.

PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149902.
PR	23-AUG-1999;	99US-0149930.
PR	25-AUG-1999;	99US-0150566.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159328.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	12.1%	Score 231	DB 21	Length 426
Best Local Similarity	30.2%	Pred. No. 8.7e-15		
Matches 84	Conservative 39	Mismatches 109	Indels 46	Gaps 11
QY	113	MESLTWNIDGDDNNLSERAGVCSYALASPDYIFLOEYIPRPYSYLKKR-----SS	166	
Db	160	ILSYVWWRREDIELN--LRMPAIGHLLQLHSPHICFOEYVPEYIDYIFRSNMKRAYSC	216	
QY	167	NYEILITGHEEGFYATIMLKSRVKLSQSEIIPFPSTKMMRNLLCVHWVSGNE-LCLMTS	225	
Db	217	SVSVDVANSRGYICMLL---SKLGYKSTSSSKSFGMSIMGRELSTAEVEYPPGKKPLPVFMTS	273	
QY	226	HLSE-----TRGHAERNNOLKMLVTLKMQOAPESATVIFAGDTNLRDREVTFCGG--	275	

[illegible]

Search completed: November 14, 2002, 10:27:24
Job time : 31.293 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 26.1584 Seconds
(without alignments)
2914.450 Million cell updates/sec

Title: US-09-697-863a-4
Perfect score: 1968
Sequence: 1 MASSSSDAEPAGPAGRAA.....DCGRPPDHGILCTLVNL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_21:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1968	100.0	370	11 Q9JUX7	Q9JUX7 mus musculus
2	1258	63.9	362	4 Q9NYY9	Q9NYY9 homo sapien
3	1258	63.9	362	4 Q9S551	Q9S551 homo sapien
4	1249	63.5	362	4 Q9NNU5	Q9NNU5 homo sapien
5	849	43.1	157	11 Q9P7N0	Q9P7N0 mus musculus
6	432.5	22.0	362	5 Q9XMG3	Q9XMG3 caenorhabdi
7	275	14.0	441	10 Q9SA95	Q9SA95 arabidopsis
8	125.5	6.4	449	5 Q9VGS4	Q9VGS4 drosophila
9	118.5	6.0	253	16 Q9S2L9	Q9S2L9 streptomyce
10	111.5	5.7	252	16 Q9ZX70	Q9ZX70 rhizobium m
11	102	5.2	265	16 Q8RUEP1	Q8RUEP1 agrobacteri
12	102	5.2	335	2 Q9RUEP1	Q9RUEP1 listeria iv
13	101.5	5.2	405	5 Q1S921	Q1S921 trypanosoma
14	101	5.1	1114	5 Q27811	Q27811 trilineus
15	100.5	5.1	502	3 Q9Y7M8	Q9Y7M8 schizosacch
16	100.5	5.1	648	2 Q8R053	Q8R053 leptospira

17	100	5.1	329	2 Q68855	Q68855 staphylococ
18	99.5	5.0	678	10 Q9C8J8	Q9C8J8 arabidopsis
19	99	5.0	327	5 Q9U2B5	Q9U2B5 caenorhabdi
20	98.5	5.0	765	16 Q8ZQ02	Q8ZQ02 salmone
21	97.5	5.0	644	4 Q9NRR6	Q9NRR6 homo sapien
22	97	4.9	447	10 Q9PFR0	Q9PFR0 arabidopsis
23	97	4.9	3571	10 Q9S127	Q9S127 arabidopsis
24	97	4.9	3574	10 Q9AUB4	Q9AUB4 arabidopsis
25	96.5	4.9	840	16 P72843	P72843 synecocyst
26	96	4.9	626	2 P77988	P77988 thermococ
27	96	4.8	633	10 Q9L145	Q9L145 oryza sativ
28	95	4.8	259	16 Q8ZRM4	Q8ZRM4 salmone
29	94.5	4.8	549	5 Q60963	Q60963 leishmania
30	94.5	4.8	765	2 Q9XCX2	Q9XCX2 salmone
31	94	4.8	1240	3 Q9P605	Q9P605 neurospora
32	94	4.8	2911	5 Q9BLV4	Q9BLV4 leishmania
33	94	4.8	3040	5 Q9GNV4	Q9GNV4 leishmania
34	93	4.7	270	16 Q910T9	Q910T9 pseudomonas
35	92.5	4.7	288	5 Q45711	Q45711 caenorhabdi
36	91.5	4.6	313	17 Q978R5	Q978R5 thermoplas
37	91.5	4.6	792	5 Q9NMV3	Q9NMV3 leishmania
38	91.5	4.6	825	11 Q912F8	Q912F8 mus musculu
39	91	4.6	4024	4 Q8WXX0	Q8WXX0 homo sapien
40	90.5	4.6	612	10 Q9LGT1	Q9LGT1 oryza sativ
41	90.5	4.6	617	12 Q83534	Q83534 measles vir
42	90.5	4.6	617	12 Q83537	Q83537 measles vir
43	90	4.6	608	16 Q9A3Q7	Q9A3Q7 caulobacter
44	90	4.6	684	12 Q9ENK9	Q9ENK9 colorado ti
45	90	4.6	691	17 Q9HPV0	Q9HPV0 halobacteri

ALIGNMENTS

RESULT 1	Q9JUX7	PRELIMINARY:	PRT:	370 AA.
ID	Q9JUX7			
AC	Q9JUX7			
DT	01-OCT-2000 (TRENBLREL. 15, Created)			
DR	01-OCT-2000 (TRENBLREL. 15, Last sequence update)			
DT	01-JUN-2002 (TRENBLREL. 21, Last annotation update)			
DE	Putative TRAF and TNF receptor associated protein.			
GN	TRAF OR TRAF.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20309820; PubMed=10764746;			
RA	Pype S., Declercq W., Ibrahim A., Michiels C., Vandenabeele P.,			
RA	Van Rietshoven A.G.I., Dewulf N., de Boer M., Vandenabeele P.,			
RT	Huybrechts D., Remacle J.E.,			
RT	"TRAF, a novel protein that associates with CD40, Tumor Necrosis			
RT	Factor (TNF) Receptor-75 and TNF Receptor-associated Factors (TRAFs),			
RT	and that inhibits Nuclear Factor-kappaB activation.",			
RL	J. Biol. Chem. 275:18586-18593(2000).			
DR	EMBL; AJ251328; CAB92971.1; -.			
DR	MGD; MGI:1860486; TrAF.			
DR	InterPro; IPR005135; Exo_endo_phos.			
DR	Pfam; PF03372; Exo_endo_phos; 1.			
KW	Receptor.			
SO	SEQUENCE 370 AA; 41033 MW; A773A8889DF5BE83 CRC64;			
Query Match	100.0%; Score 1968; DB 11; Length 370;			
Best Local Similarity	100.0%; Pred. No. 3.2e-168;			
Matches 370; Conservative 0; Mismatches 0; Indels 0; Caps 0;				
QY	1 MASSSSDAEPAGPAGRAA...PEAAQAEDRYKRRRLQCGFALVGGCDPTWPSVIRE 60			
DB	1 MASSSSDAEPAGPAGRAA...PEAAQAEDRYKRRRLQCGFALVGGCDPTWPSVIRE 60			
QY	61 NDWQKALSAVFELPNDQGWPRPPTSFKSEAVYDLTNDANDTTLLEASPSGTPLD 120			

Db	61	NDMOTQKLSAYFELPEPDGSGMPQPPTRSKSEAYVLDINEDADDTILLASPSGPLED	1200
QY	121	SSITSIFITMWNIDGLDGNLPERAGVSCSLATSPDVVFLQEVLPYCAVLAKKRAASYTI	1800
Db	121	SSITSIFITMWNIDGLDGNLPERAGVSCSLATSPDVVFLQEVLPYCAVLAKKRAASYTI	1800
QY	181	ITGNEEGYFTAILLKKGVRKKSQSEIIPFPNTKMMRLCCVNSVLGNEFCMLMTSHLEST	2400
Db	181	ITGNEEGYFTAILLKKGVRKKSQSEIIPFPNTKMMRLCCVNSVLGNEFCMLMTSHLEST	2400
QY	241	REHSAERIROLKTYLQKMQEAPDSTYIIFAGDTRLROEYVTKCGGLPDNVEFDMAEFLGKP	3000
Db	241	REHSAERIROLKTYLQKMQEAPDSTYIIFAGDTRLROEYVTKCGGLPDNVEFDMAEFLGKP	3000
QY	301	KHCQYTMDTKANNNLRIPAAVKHNFDRIFPRAEEGHILIPSLDVLGPEKLDGCRFPESDHW	3600
Db	301	KHCQYTMDTKANNNLRIPAAVKHNFDRIFPRAEEGHILIPSLDVLGPEKLDGCRFPESDHW	3600
QY	361	GLLCTLNVYL 370	
Db	361	GLLCTLNVYL 370	

	RESULT 2		
Q9NRY9	Q9NRY9	PRELIMINARY;	PRT; 362 AA.
AC	Q9NRY9;		
DT	01-OCT-2000 (TREMBLrel. 15, Created)		
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)		
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)		
DE	AD022 protein.		
GN	AD022.		
OS	Homo sapiens (Human)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=ADRENAL GLAND.		
RA	Song H., Gao G., Peng Y., Ren S., Chen Z., Han Z.;		
RT	"A novel gene expressed in human adrenal gland.";		
RL	Submitted (JAN-2000) to the EMBL/Genbank/DBJ databases.		
DR	EMBL: AFZ23463; AAF614.1; -.		
DR	InterPro: IPR005135; Exo_endo_phos.		
DR	Pfam: PF03372; Exo_endo_phos; 1.		
DR	SEQUENCE 362 AA; 40956 MW; D6D607D5E21C27E4 CRC64;		

Query Match	63.9%;	Score 1258;	DB 4;	Length 362;
Best Local Similarity	68.4%;	Pred. No. 1.6e-104;		
Matches 240;	Conservative 43;	Mismatches 64;	Indels 4;	Gaps 3;

Qy	24	EAQOAE--EGRVARRLOGLFALGAGGDDPTMPVSYLRENDMOTOKALASAELEENOGW	82
Dd	12	EAAEEGGEYEVYKRRLLCYVRASVASCDAAVAOCLFAEMDEMERALNSFEPPVESAL	71
Qy	83	PROPTSFKSEAYVDLTJNEBANDNTTILEASPS--GTELEDSSTISEITNNIDGLDGCNPE	144
Dd	72	ERRPETISEPKTYVLTJNEETTTDSTSKISPSBDTQENGSMFSLITNNIDGLDNLNLE	131
Qy	142	RANGVSCALASPPVNVLOEVIIPRYCALAKRRAAYTTITGNEGCTYATILLKKGRYF	201
Dd	132	RANGVSCALASPPVIFLOEVIIPRYCYLYTKRSSWEIITHEGCTYATILLKRSYKL	191
Qy	202	KSGEILIPENTKMBNNLLCVNLSLGNELFCMLMTHLESTREHSAIERIOLTYTGKKMOEA	261
Dd	192	KSGEILIPSPSTKMBNNLLCVHVNSGNELCLMTHSHLESTRGAABRMQNLKMYLKKMOEA	251
Qy	262	PDSITYIIFNGDNLNLDQEVIKCGSLPPNVVDANBEFGAKRKHQCYTWTKRANNULRPAY	321
Dd	252	PESATYIIFNGDNLNLDREVYTRCGGLPNNIYDVWFEIGAKRKHQCYTWDMQNNNSGITPAAC	311
Qy	322	KHFDRIEFR--AEEGHILIPQSLDVLVEKLDGRRPSHHMGLTTLNVVL	370

```
Db      312 KLRFDRIFFRAAEERGHII PRSLDLGLEKLDGCRFPSPDHWGILCNLDIIL 3622
```

RESULT 3
095551
ID 095551 PRELIMINARY: PRT: 362 AA.

DT 01-MAY-1999 (TREMblrel. 10, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE D3J0M3.2 (TYRaf and TNF receptor associated protein) (ETSI-associated
DE protein) (2) (Hypothetical 40.9 kDa protein).
GN D3J0M3.3 OR TRRAP OR EAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RP Phillips S.;
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.

RP MEDLINE=20309820; PubMed=10766746;
 RX Pye S., Deelercq W., Ibrahim A., Michiels C.,
 RA Van Riettschoten A.G.I., Dewulf N., de Boer M., Vandenaebelle P.,
 RA Huybroeck D., Remacle J.E.;
 RT "TRAP, a novel protein that associates with CD40, Tumor Necrosis
 RT Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs)
 RT and that inhibits Nuclear Factor-kappaB activation.";
 RL J. Biol. Chem. 275:18586-18593(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RP Li R., Pei H., Papas T.S.;
 RT "EAP2, a Novel Protein Interacting with ETS1.";
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RP TISSUE=SKIN;
 RA Strausberg R.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 EMBL AL031775; CAA21141.1; -
 DR EMBL; AJ269473; CAB92966.1; -
 EMBL; AF201687; AAG35600.1; -
 DR EMBL; BC017553; AAH17553.1; -
 DR InterPro: IPR005135; Exo_endo.phos.
 DR Pfam: PF03572; Exo_endo.phos; 1.
 RP Receptor; Hypothetical protein.
 KW SEQUENCE 362 AA; 40929 MW; 37892E125DB64410 CRC64;
 QM

Query Match	63.9%;	Score 1258;	DB 4;	Length 362;
Best Local Similarity	68.4%;	Pred. No. 1.6e-104;		
Matches 240;	Conservative 43;	Mismatches 64;	Indels 4;	Gaps 3;

[illegible]

[illegible]

RESULT 4

ID	Q9NUR5	PRELIMINARY;	PR;	362	AA.
AC	Q9NUR5;				
DT	01-OCT-2000 (Tremblay1. 15, Created)				
DT	01-OCT-2000 (Tremblay1. 15, Last sequence update)				
DT	01-JUN-2002 (Tremblay1. 21, Last annotation update)				
DE	CCNA FL11306 fls, clone PLACE1010031.				
OS	Homo sapiens (Human)				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
OX	NCBI_TaxID=9606;				

RP SEQUENCE FROM N.A.
RC TISSUE=PLACENTA;
RA Isegal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatauma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi
RA Watanabe S., Kimura K., Murakami K., Ishi S., Kawai Y., Saito
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
RA Ninomiya K., Iwayanagi T.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK002168; BAA92119.1; -
DR InterPro: IPR005135; Exo_endo.phos.
DR Ptm; PF03372; Exo_endo.phos; 1.
SQ SEQUENCE 362 AA; 40869 MW; B81702C93BCB62E3 CRC64;

Query Match	63.58;	Score 1249;	DB 4;	Length 362;
Best Local Similarity	68.18;	Pred. No. 1e-103;		
Matches 239; Conservative	43;	Mismatches 65;	Indels 4;	Gaps 3

Qy 24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTMVPVSURENDMOOKALSAYFELPENOGW 82
||| : | ||| : || : ||| : ||| : ||| :
Db 12 EAEEEGEDEVKKRRLCVERASVASCDPAVAOCFLAENDMEMRALNSCEPPEESAL 71

```

QY      83  PRQPTSEKSEAYVDLTNEDANDTTILEASPS-CGPLEDDSTISFIWNIDGLDGNLPE 141
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      72  ERPRPTESEPKTYVDLTNEETNDSTTSKISPSPEPTQENGSHFLITWNIDGLDNLNLS 131

```

Dh 132 RARGVCSYALYSPTVFLOEYIPDPVSYLKKRSSNVEITGHEEGYFAIMKKSRVKL 191

QY 142 RARGVCSCLAYSPDVFLOEYIPPYCAVLKKRASYYIINGNEGFTAILKKGRVKE 201

QY 202 KSOEIIFFPNTKMMNLLCVANSVLGNEFCMLTSHLESTREHSAERIRQLKTVLCKMOEA 261
|||||::||| ||||| :||| |||
Db 192 KSOETIPDECTKMANTLCVHNUSGNFCTMTSHLESTRGDAAPPMNOI KMVLKKMOEA 251

```

0y      262 PDSTVIFAGDTNLRDQEVKGGGLPDNVFDAWELGKPKHCQYTWDTKANNNLRIPAY 321
        | : | | | | | | | | | | : | | | | | | | | | | | | | | | | | | | | | |
0c      263 DPGATUTPRGDMMRDPDEWEPBOOTDNNTVDUWEETCKDKHCVOWEEMTDOMCNTCTAAG 311

```

[illegible]

RESULT 5	
Q9D7N0	
ID Q9D7N0	PRELIMINARY; PRT; 157 AA

DT 01-JUN-2001 (TREMblurel. 17, Created)
DT 01-JUN-2001 (TREMblurel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblurel. 21, Last annotation update)
DE Trif and Tnf receptor associated protein.

GN TTPAP.
OS Mus musculus (Mouse).
OC Euryarchaeta: Metazoa: Chordata: Vertebrata: Euteleostomi:
OC Mammalia: Eutheria: Rodentia: Sciurognathi: Muridae: Mus.
OX NCBI_TaxID=10090;
NN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=TONGUE;
RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinaeawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.
RA Arkawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishii K., Kiyosawa H., Kondo S., Yamanaka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Flitschmann W., Gaasterland T., Glasi C., King B., Koehisa H.,
RA Kuehl P., Lewis S., Matsuo T., Nikaido I., Pesole G., Quackenbush J.,
RA Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Canciani P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald C., Rodriguez I., Sakamoto N.,
RA Saeki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weltz C., Whitaker C., Wilming L.,
RA Wyszyna-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA Hayashizaki Y.,

RT Functional annotation of a full-length mouse CUNA cDNA
 RL Nature 409:685-690(2001).
 DR EMBL: AK009089; BAB26063.1; -.
 DR MGI: 18600486; Tetrap.
 DR InterPro: IPR005135; Exo_endo_phos.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 SO SEQUENCE 157 AA; 17876 MW; 5920850802FAAE84 CRC64;

Query Match	43.1%	Score 849;	DB 11;	Length 157;
Best Local Similarity	100.0%	Pred. No. 2.4e-68;		
Matches 157; Conservative	0;	Mismatches 0;	Indels 0;	Gaps 0;

[illegible]

QY 274 NLRDQEVICGGLPDNYFDAMEFLGPKHCQYTWDTKANNNLRIPAAKYHREDRIEFPRAE 333
|||||
Db 61 NLRDQEVICGGLPDNYFDAMEFLGPKHCQYTWDTKANNNLRIPAAKYHREDRIEFPRAE 120
|||||

QY 334 EGHIPQSLDLVGLKLDGGRFSDHWGLCTLNVL 370
 |||||
 DB 121 EGHIPQSLDLVGLKLDGGRFSDHWGLCTLNVL 157

RESULT 6

ID	Q9XW63	PRELIMINARY;	PRI;	302 AA.
AC	Q9XW63;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		

DT	01-JUN-2002 (TREMBlrel. 21, last annotation update)
DE	Y63D3A.4 protein.
GN	Y63D3A.4.

0C Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabditoidea
0C Rhabdilitidae; Peloderiinae; Caenorhabditis.
0X NCBI_TaxID=6239;

RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.

RP	SEQUENCE FROM N.A.
RX	MEDLINE=99069613; PubMed=9851916;
RA	none;

```

RT "Genome sequence of the nematode C.elegans: A platform for
RT Investigating biology."
RL Science 282:2012-2018(1998).
DR EMBL: AL033652: CAA21707.1:..
DR InterPro: IPR005135: Exo_endo_phos.
DR Pfam: PF03372: Exo_endo_phos; 1.
SQ SEQUENCE 362 AA: 40893 MW: 78954750486652 CRC64;

Query Match 22.0%; Score 432.5; DB 5: Length 362;
Best local similarity 33.5%; Pred. No. 1.9e-30;
Matches 113; Conservative 55; Mismatches 126; Indels 43; Gaps 11;

QY 43 FALVGGDPTWVPSVLRNDMOTOKALSAFELPNDGQWPRQPTSKSAIVDLNED 102
DB FAITRTDEAFQSLQDVDMDLKALDYF-----GSEAF-----E 89
QY 103 ANFTTLEAS---PSGPTL---ED--SSTIFITWNIDGLDCNLPRRAGVSCIALY 153
DB ARSAAVMGASSSSMASSGAAVMTAEDLKGEVSMNMINDGLDRSLTRMAVAHIVKNV 149
QY 154 SDPVFLQEVIPYCAVLYKKRAASTYITGNEE-GTFALLKGRVKSOEIIIPPT 212
DB NPDLLEQEVWDRLAPIDIKLSLYKYYNKGCOYTAIIYVK-MFVKEKHVYIHQNS 208
QY 213 KMRNLICVNVSLGNEFCLMTSHLESTREHSAERIRQLKTVLGKMOE---APDSTYIF 269
DB GMRTRTQILEGSLIGLKVFLNLTHLESTREHHPQRCAPFCMDKVRRLIAQNGALVYF 268
QY 270 AGDTNLKDEVIKCGLPNVFDAMEFLGKPRHCOYTWDTKANNNLRIPAAVYKREDRIF 329
DB GGLNLRDEVSER---VPGVKDAMEAGSDMKRTFTWDTFKNDKOGFHGAKMRFRLY 325
QY 330 FRAEEGHLIPQSLDLVGLKEU-DCGRFSPDHGCLCT 365
DB 326 W---SGPLDKVKRTLEGQRIRISCLCFPSDHAINNT 359

RESULT 7
Q9SA95 PRELIMINARY: PRT; 441 AA.
AC O9SA95;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE F25C20.3.
GN F25C20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremetskaia I., Lueros J., Ngan L., Conway A.B.,
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shin P.,
RA Walker M., Davis R.W., Becker J.R., Peterspil N.A., Theologis A.;
RT Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.
RL Submitted (APR-1999) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CV. COLUMBIA;
RA Theologis A.;
RT Submitted (MAY-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AC007296: AAD30241.1:..
DR InterPro: IPR001969: Asprotease.site.
DR InterPro: IPR005135: Asprotease.site.
DR InterPro: IPR001876: znf_RangDP.
DR Pfam: PF03372: Exo_endo_phos; 1.
DR SMART: SM00547: znf_RBZ: 2.
DR PROSITE: PS00141: ASP_PROTEASE; UNKNOWN_1.

```

```

DR PROSITE: PS01358; ZF_RANBP2_1; UNKNOWN_1.
SQ SEQUENCE 441 AA: 48995 MW: 6582FPAAD27BD0CA8 CRC64;

Query Match 14.0%; Score 275; DB 10: Length 441;
Best local similarity 31.4%; Pred. No. 3.4e-16;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASPSGTPLEDSSTISFTIWNIDGLDCNLPERAGVSCIALYSPDVVFLQEVIPYCAV 170
DB ASDSGTPLCLKILSYNVWFREDLE---LNLRMRAIGHLIQHSPLILCFQEVPEYDI 219
QY 171 LKKR-----AASYITITGNEGYFTALLKGRVKEFSQELIPPTKMMRNLCVNS 224
DB FRSKNMKNKAYSCSVSDVAVSNGRYGICMLSKGVKSSKSK---FGNIMKRELSTAEVE 276
QY 225 LGGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGKMOEAPDSTYIPAGDTNL 275
DB VPGKRPFLVATSHLESPPCPGPKMDQMFSSREEQAKFAIELR---PNANVIFGDMNV 333
QY 276 RQEVYKCG---LDNVFDAMEFLGKPRHCOYTWDTKANNNLRIPAAVYKRFPRIFRA 332
DB 334 CD---KLDGKRPPLDPKVDVWEVL-KPQDLGFTYDTRANPMLSGNRLQRLDLICRL 388
QY 333 EESGHLIPQSLDLVG-----LEKIDCGRPSDHGCLCTLN 367
DB 389 DDYKL--GGIEWVGKEAIPGLSYVEKEKVRGDIKLELPVLPSPHFCLLYTLS 439

RESULT 8
Q9VGS4 PRELIMINARY: PRT; 449 AA.
AC Q9VGS4;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE CG4796 protein.
GN CG4796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephyroidae; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RX MEDLINE=20196006; PubMed=10731132.
RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Adayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Burks K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclob J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,

```



```

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021:
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorheeler F.J., Hernandez-Lucas I., Becker A., Couzy J.,
RA "The complete sequence of the 1,663-Kb pSymb megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
DR EMBL: AF603642; CAC48492.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
KW Plasmid; Hypothetical protein; Complete proteome.
SQ SEQUENCE 252 AA; 27376 MW; ABB6F28C8A8316AE CRC64;

Query Match
Best Local Similarity 21.8%; Score 111.5; DB 16; Length 252;
Matches 62; Conservative 36; Mismatches 102; Indels 85; Gaps 11;

QY 124 ISFTIWNIDGICNLPERARGVCSCLALYSPDVVFLQEV-----IPPYCAVL 171
Db 12 IRLTYNVHSCFTGDRDRLPARIAVIAECQPDVIALQEDVNGRARTGSDAAMIAIHL 71
QY 172 KKRAASYITITGNESGYFAILLKGRVYKKSQEIIPPNTRMMRNLLCVNVSLSGNEFC 231
Db 72 NMEAEFHPALHLEDEKYGDAVLT---ALPMRLIKAAPLPSSEPGALWEIDVAAVKLQ 128
QY 232 LMTSHLESTRHSABRIQLKTVLGR---MQAPSTYIYFAGDIN----- 274
Db 129 VYATHLGL---RGAERLRKATALTALGPGWLGMAQDAHVLAIGDANATGRSTAYRLARQ 185
QY 275 LRDOEVYIKGGPLPDVNFDAWMEFLGKPKHCQYQWTDTKANNLRIIPAAYKHRRPDRIFRAEE 334
Db 186 LSDAQLL-TGVAPRPPTFPSS-----RLPLD---RIDVH----- 213
QY 335 GHILPOSIDLTV--GLEKLDLC-----GRPSDHMGILLCTLNVL 370
Db 214 -----LVGKGIEVASCVRHGSTLARSASDHLPLLAELDVVM 249

RESULT 11
OBUPE1 PRELIMINARY; PRT; 265 AA.
AC 08UPE1;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Exonuclease III.
GN ATU1714 OR AGR_C_3151.
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Rhizobium.
OX NCBI_TaxID=176299;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=21608550; PubMed=11743193;
RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyee D. Sr.,
RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
RA Kulyavcin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimachak C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Kraspan W., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tiney S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Nester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RC58."
RL Science 294:2317-2323(2001).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21608551; PubMed=11743194;

```

```

RA Goodner B., Hinkle G., Gatlung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Gao Y., Askenazi M., Halling C., Mullin L.,
RA Houmlet K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
RA Cielo C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL: AE009128; AAU42714.1; -.
DR EMBL: AE008093; AAK87487.1; -.
KW Complete proteome.
SQ SEQUENCE 265 AA; 29818 MW; 361FA77E3FEA979 CRC64;

Query Match
Best Local Similarity 23.5%; Score 102; DB 16; Length 265;
Matches 67; Conservative 34; Mismatches 94; Indels 90; Gaps 16;

QY 128 TWINIDGICNLPERARGVCSCLALYSPDVVFLQEV--IPPYCAVLKKRAASYITITGNE 185
Db 5 TWINNGVKA-----RIENLCOMLKDSSPDIVCQDEIKSVDEGPRFLEALGHYHIEHQ 59
QY 186 EGYFTAILLKGRVYKKSQEI--IPPNTRMMRNLLCVNVSLSGNEFCIMTSHLESTRE 242
Db 60 KGFNGVALLS---KVPDEYVNRGLPGDDADEQARFIEGVFSEGAIRVCSLYLPGNP 115
QY 243 HS-----AERIRQLKTVYKQGEAPDSTTVYFAGDTNLRDOEVYIKGGPLDNVF 291
Db 116 PDDEPVKYRYKLAMERLRRAEDRLAME---PLTIADYV-----IPE-PP 159
QY 292 D-----AME-----FLGKPK-----HCQYT-----WDTKA-----NNNL 315
Db 160 DCHDPRWEGDGLFLPRTKRAAFKLEMLGFTDAARATTDAAGLXSRNDYAGAMPKNNKI 219
QY 316 RIPAAYKHRRPDRIFFEAEBSHLIPQSLDLVLEK-LDCGRFPSPDH 359
Db 220 RI-----DHMLSAEAA-----DRLOSVEIEKHVAMKPSDH 252

RESULT 12
O9RLV9 PRELIMINARY; PRT; 335 AA.
AC 09RLV9;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Sphingomyelinase-c.
OS Listeria ivanovii.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillales;
OC Listeriaceae; Listeria.
OX NCBI_TaxID=1638;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19119;
RX MEDLINE=99348382; PubMed=10417642;
RA Gonzalez-Zorn B., Dominguez-Bernal G., Suarez M., Ripio M.T., Vega Y.,
RA Novella S., Vazquez-Boland J.A.;
RT "The smcL gene of Listeria ivanovii encodes a sphingomyelinase C that
RT mediates bacterial escape from the phagocytic vacuole."
RL Mol. Microbiol. 33:510-523(1999).
DR EMBL: Y09477; CAA70683.2; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 335 AA; 38455 MW; 46F3985E8A8434D7 CRC64;

Query Match
Best Local Similarity 21.8%; Score 102; DB 2; Length 335;
Matches 51; Conservative 32; Mismatches 83; Indels 68; Gaps 12;

QY 156 DVVFLQEVIPPYCAV-----LKKRAASYITITGNE-----EGYFTAILLKGRVYK-- 201
Db 82 DVVILNEAFDTSASHRLNNLRMPYHQTPVIGRSKHGMDKTEGVSFNALBDGCVAAVVS 141

```

QY	202	-----KSOELIP-----	FPNTKMRNLICVAVSLGNGDFCLMTSHL-----	237
Db	142	OMPIYEKSOHIFQRCGGADRLSNKGAAYVTKIKN-----	-----GKPYHLITHTTPADSL	192
QY	238	---ESTREHSAERIRQLKTLVGLKMGADPSTVIFACPTNLRDQEVIKCGPLDWNFPAW		294
Db	193	ISKDTSRAIRABOWOEIQEFIAK-KNIPRDEILIFGGDLN-----	VNNG--TDEHDMF	243
QY	295	EEIGKPKHCQY-----	TWDTKANNNLR--IPAAVYHREDRIEFPRAEGLTIQOS	341
Db	244	KLLNVSPANFNGQMATWDPPTNSMLKESYKPAAEYLDIYI--	VENGHARPHS	295
RESULT 13				
ID	015921	PRELIMINARY:	PRT:	405 AA.
AC	015921.			
DT	01-JAN-1998	(TREMBLrel, 05, Created)		
DT	01-JAN-1998	(TREMBLrel, 05, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel, 21, Last annotation update)		
DE	AP-endonuclease.			
GN	TCAP.			
OS	Trypanosoma cruzi.			
OC	Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.			
OX	NCBI_TaxID=56933;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-Y;			
RX	MEDLINE=99108141; Pubmed=9889272;			
RA	Perez J., Gallego C., Bernier-Villamor V., Camacho A.,			
RA	Gonzalez-Pachonoka D., Ruiz-Perez L.M.;			
RT	"Auricularia/apurimilidic endonuclease genes from the trypanosomatidae			
RT	leishmania major and Trypanosoma cruzi confer resistance to oxidizing			
RT	agents in DNA repair-deficient Escherichia coli."			
RL	Nucleic Acids Res. 27:771-777(1999).			
DR	EMBL; U92486; AAD1456.1; -.			
DR	HSSP; P27695; 159N.			
DR	InterPro; IPR004442; ExoDNase_III.			
DR	InterPro; IPR004808; ExoIII_xth.			
DR	InterPro; IPR005135; Exo_endo_phos.			
DR	Pfam; PF03372; Exo_endo_phos; 1.			
DR	TIGRFAMS; TIGR00195; exoDNase_III; 1.			
DR	TIGRFAMS; TIGR00633; xth; 1.			
KW	Endonuclease			
SO	SEQUENCE 405 AA; 45193 MW; 3BAFC182DB8FC6 CRC64;			
Query Match 5.2%; Score 101.5; DB 5; Length 405;				
Best local Similarity 20.7%; Pred. No. 1.1;				
Matches 84; Conservative 37; Mismatches 119; Indels 165; Gaps 22.				
QY	86	PPTSKSAVYDITNEDANDTITILEASTSGIPLEDS-----		122
Db	44	PPSRSLNSAGAFAVSPNRPPLAAYLTPP--PSDDTRKTEKDIMSQVEFORRTAKDF		100
QY	123	---TISTITNIDIDGCNLPERRAGVSCSLATSPDVVFLQEV-----	I	164
Db	101	DSKHLKRTITNNVAGLRGL-LKDDQALQRLLEEGRPALCQLQETKLNPDQNEKLGCV		159
QY	165	PPY-----C-----AYLKKRAAS-YTIT-		187
Db	160	PGYRFVDVHCRAKKGSGSTRYIKNTAAEMKTVYKGFDTLKSPODVGSHSGDEGRVL		219
QY	188	---YFAILLKKGRAVFKSOELIP--PNTKMRNL-----CVNVSLOGNFCMLTSHLESTR		241
Db	220	TTYFT--QGSGSEFFALALAVNTYIPNMGSLERIPYRCQKFDL-----		261
QY	242	EHSARIRQLKTLVGLKMGADPSTVIFACPTNLRDQEV-----	VIKC	283
Db	262	-----RIHQHCTIGRSCNHDKEEDASLIGFTIAGDGLNAERYDYRTFASYSKAMQKC		316
QY	284	GGL-PD-----NVDNAEFFLCKPKHCQYT-WDTKANNNLRIPAAVYKHRFDRIE		329
Db	317	SGFPEERASRETLRVANAADTFALPKPAAYVTFPSARINGRAR--GLGMDLDYFV		373

Qy	330	FAAEESHILPQSLD.VGLEKLDGCRF----	SDHMGILCTLWVVL	370
			:	:
Db	374	VSA-----	ALARHVDCFTPHVMGSDH----	CPLOMWL 403
RESULT 14				
ID	027811	PRELIMINARY;	PRT;	1114 AA.
AC	027811;			
DT	01-NOV-1996	(TREMBLrel. 01, Created)		
DT	01-NOV-1996	(TREMBLrel. 01, Last sequence update)		
DT	01-MAR-2002	(TREMBLrel. 20, Last annotation update)		
DE	Dynein heavy chain isoform 7A (EC 3.6.1.3) (Fragment).			
GN	DYH1A.			
OS	Trinectes gratilla (Hawaiian sea urchin).			
OC	Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;			
OC	Echinoidae; Euechinoidae; Echinacea; Temnopleuroidea; Toxopneustidae;			
OC	Trinectes.			
OX	NCBI_TaxId=7673;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=EMBRYO;			
RA	MEDLINE=94243035; PubMed=8186465;			
RA	Gibbons B.H., Asai D.J., Tang W.T., Hays T.S., Gibbons I.R.;			
RT	"Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchins."			
RL	Mol. Biol. Cell 5:57-70(1994).			
EMBL	U03978; AA63592.1; -			
DR	InterPro: IPR002048; EF-hand.			
DR	Pfam: PF00036; ehand; 1.			
DR	PROSITE: PS00018; EF_HAND; UNKNOWN_1.			
KV	Hydrolase.			
FT	NON_TER	1	1	
FT	NON_TER	1114	1114	
SO	SEQUENCE	1114 AA;	125999 MW;	BCF09927EF5688BC CRC64;
Query Match				
	Best Local Similarity	5.1%;	Score 101;	DB 5; Length 1114;
	Matches	53; Conservative	36; Mismatches	87; Indels 72; Gaps 11;
Qy	12	PAGPAGRAASAEPAQAQEDRYKRRRLQCLGALVGGCDPTWVPSVLRENDWQTKALSA	71	
Db	39	PGPAGTGT-----ETTKDLAKAVAKKQCVVENCSDGLD-----YIALGK	78	
Qy	72	YVELPNDGWRPQPTSPKSEAYVDL-----TIEDANDTTLFASPSG	115	
Db	79	PFKGLASGCAW-----SCDFENRIDLEVLVVAOOIILTIQGINAGA-DTLFFE----	128	
Qy	116	PLEDSSITS-FITWNIIDLGDCNIPERARGVCSGCIATSPDVVFLOEYIPPYCAYLKKR	174	
Db	129	TEIKLDPTCSVFTTMNPGYAGRSIDLPLNKAIFRVAAMVPYALISELVLSGCTIKR	188	
Qy	175	AASYITIT-----GNEEGY-----FTAILLKKGVKKK-----SQEIIPEPTTKM	215	
Db	189	PLSVKIVATYRLCSQSLSSQHHYDYGMRVAKSVLTAAAGNLKIKYPEEDDL-----ML	242	
Qy	216	RMLLCVNV	223	
Db	243	RSINDVNL	250	
RESULT 15				
ID	Q9Y7M8	PRELIMINARY;	PRT;	502 AA.
AC	Q9Y7M8;			
DT	01-NOV-1999	(TREMBLrel. 12, Created)		
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Hypothetical 57.7 Kda protein.			
GN	SPBC986.11C.			
OS	Schizosaccharomyces pombe (fission yeast).			
OC	Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;			

OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=972H-;
RA Wood V., Rajandream M.A., Barrell B.G., Vojackert G.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL049769; CAB42372.1; -
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF0372; Exo_endo_phos.1.
KW Hypothetical protein.
SQ SEQUENCE 502 AA; 57650 MW; 54DFE2E7613B1A86 CRC64;

Query Match 5.1%; Score 100.5; DB 3; Length 502;
Best Local Similarity 18.1%; Pred. No. 1.9; Indels 183; Gaps 18;
Matches 88; Conservative 61; Mismatches 154;

QY 62 DWOTOKALSAYFE-----LPENDGWPROPPTSEAYVDLTNEDANDTTILEAS 112
DB 22 DYRLSKLAISFPDQKNKMAADSIPIVKKKSAKGP-SFVTPEYIEKQKLEKMAKAA 80
QY 113 -----PSTPLE-----DSSTISFTIWNIDGLGCNL----- 139
DB 81 RKPIAPPSNAPPEFNTDFIKREMLSPNYAPETEKSALDITIMYNY-LAQTNIRSM 138
QY 140 -----PERARGVCSCLALYSPDVFLQEVIPRYCAVLKRR--AASYTI--ITGN 184
DB 139 PFHSEALKWKNRSRLMNLTYISPTIGCMQEVDAEFVPMFYKKLGLGLEYLHFIKGE 198
QY 185 EEGYFTAILLKGRVYKFSOEIIPE-----PNTKMMNLLCY---NVSIGNEFCIM 233
DB 199 GKTGIMIFMKSSLFK-KVQDLTIYDDHDELPGRMNTKNIGCCVRLERVDPSRGLEFLA 257
QY 234 TSHLESTREHSERIRQ---LKTIVGKMQEAPDSTTVIFAGDTN-----LRQDEV 280
DB 258 TTHLFWHRYGSYERLRQGAIIIVKEVYKMAQSHPSWPVFIAGDFNTEPDTNFPALTTRPL 317
QY 281 IKCGGLPNDVFDAMEFL-GKPRHCQYTWDTKANNN----- 314
DB 318 SICQATDIIERSMNYVFGESLEKKNASTKTENDSNEDDKECCOSSSTSSVPESTASTP 377
QY 315 -----LRIPAAKKHPRRIFF---RAE 333
DB 378 KKRILHVQNDYVPHYRSFYQHQEONPVLFLSYSGYKLVHPENAKNTPEDHPAFTNMAHAY 437
QY 334 EGH-----IPQSLDLVGLKLDGGRFPDHWGL 363
DB 438 OGHLDIYFVMNRDTSLOTPENQVVEGIKLALLRVLPSEKKEAPELE-GRYPSDHYALM 496
QY 364 CTLVNV 369
DB 497 ANQIV 502

Search completed: November 14, 2002, 10:25:17
Job time : 30.1584 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 25.5928 Seconds

(without alignments)
2914.450 Million cell updates/sec

Title: US-09-697-863A-2

Perfect score: 1902
Sequence: 1 MELGSCLEGGREAREEGEP.....DCGRFSDHMLCNDLIL 362

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_prodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1902	100.0	362	4	095551
2	1897	99.7	362	4	09NY9
3	1893	99.5	362	4	09NUK5
4	1258	66.1	370	11	09JUX7
5	675	35.5	157	11	09D7N0
6	420	22.1	362	5	09XW3
7	231	12.1	441	10	09SA95
8	127.5	6.7	449	5	09VGS4
9	122.5	6.4	361	5	08T2I4
10	113	5.9	502	3	09Y7M8
11	111.5	5.9	252	16	092X70
12	108.5	5.7	720	10	09LS07
13	108	5.7	253	16	09S2L9
14	107.5	5.7	863	15	08UNX2
15	106	5.6	267	16	09CMW2
16	106	5.6	1962	11	09WUF3

17	105.5	5.5	667	11	08VCU0	08VCU0 mus musculu
18	105	5.5	1374	10	022148	022148 arabidopsis
19	104.5	5.5	281	16	08YPH2	08YPH2 anabena sp
20	103.5	5.4	256	16	08YAA1	08YAA1 listeria mo
21	103	5.4	288	5	045711	045711 caenorhabdi
22	103	5.4	1275	4	09Y5K0	09Y5K0 homo sapien
23	103	5.4	1275	4	000366	000366 homo sapien
24	102	5.4	265	16	08UEP1	08UEP1 agrobacteri
25	102	5.4	1275	4	000375	000375 homo sapien
26	101	5.3	405	5	015921	015921 trypanosoma
27	101	5.3	1275	4	015604	015604 homo sapien
28	101	5.3	1275	4	012881	012881 homo sapien
29	101	5.3	1275	4	000368	000368 homo sapien
30	101	5.3	1275	4	000378	000378 homo sapien
31	101	5.3	1275	4	015606	015606 homo sapien
32	101	5.3	1275	4	08TE30	08TE30 homo sapien
33	100.5	5.3	695	2	085672	085672 clostridium
34	100.5	5.3	816	9	08SDP5	08SDP5 pseudomonas
35	100	5.3	458	10	091VF3	091VF3 arabidopsis
36	100	5.3	1261	4	015463	015463 homo sapien
37	100	5.3	1275	4	000370	000370 homo sapien
38	100	5.3	1275	4	09UN80	09UN80 homo sapien
39	99.5	5.2	256	16	08XLI6	08XLI6 clostridium
40	98.5	5.2	283	5	09BKP5	09BKP5 dictyosteli
41	98	5.2	703	5	08SW87	08SW87 encephalico
42	98	5.2	1275	4	000360	000360 homo sapien
43	98	5.2	1275	4	000362	000362 homo sapien
44	98	5.2	3844	5	094648	094648 plasmodium
45	97.5	5.1	1001	11	09JMC1	09JMC1 rattus norv

ALIGNMENTS

RESULT 1
095551 PRELIMINARY; PRT; 362 AA.
ID 095551;
AC 095551;
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE D3J0M3.3 (TRAF and TNF receptor associated protein) (ENSL-associated
protein 2) (Hypothetical 40.9 kDa protein).
GN D3J0M3.3 OR TRAF OR EAP2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Phillips S.;
RL submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309820; PubMed=10764746;
RA Pyke S., Declercq W., Ibrahim A., Michiels C.,
RA Van Rietchooten A.G.I., Dewulf N., de Boer M., Vandenabeele P.,
RA Huybrecock D., Remacle J.E.;
RT "TRAF, a novel protein that associates with CD40, Tumor Necrosis
Factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs),
and that inhibits Nuclear Factor-kappaB activation.";
J. Biol. Chem. 275:18586-18593(2000).
RN [3]
RP SEQUENCE FROM N.A.
RA Li R., Pei H., Papas T.S.;
RT "EAP2, a Novel Protein Interacting with ENSL.";
Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC TISSUE=SKIN;
RA Strausberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL031775; CA21141.1; -.

DR EMBL: AJ269473; CAB92966.1; -
 DR EMBL: AF201687; AAG35600.1; -
 DR EMBL: BC017553; AAH17553.1; -
 DR InterPro: IPR005135; Exo_endo_phos.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 KW Receptor; Hypothetical protein.
 SQ SEQUENCE 362 AA; 40929 MW; 37892E125DB64410 CRC64;

Query Match 100.0%; Score 1902; DB 4; Length 362;
 Best Local Similarity 100.0%; Pred. No. 2e-150;
 Matches 362; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60
 Db 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60
 QY 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 Db 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 QY 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 Db 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 QY 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 Db 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 QY 241 LKWLKKMQAPESATVIFAGDTNLRDEVTRCGGLPNNIYDVWEFLGKPKHCQYTWTDQ 300
 Db 241 LKWLKKMQAPESATVIFAGDTNLRDEVTRCGGLPNNIYDVWEFLGKPKHCQYTWTDQ 300
 QY 301 MNSNLGITTAACKLRFDRIFFRAAABEGHIIIPRSLDLGLEKLDGGRFPDHWGLCLNDI 360
 Db 301 MNSNLGITTAACKLRFDRIFFRAAABEGHIIIPRSLDLGLEKLDGGRFPDHWGLCLNDI 360
 QY 361 IL 362
 Db 361 IL 362

RESULT 2

Q9N9Y9 PRELIMINARY; PRT; 362 AA.
 AC Q9N9Y9;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE AD022 protein.
 GN AD022.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=ADRENAL GLAND;
 RA Song H., Gao G., Peng Y., Ren S., Chen Z., Han Z.;
 RT "A novel gene expressed in human adrenal gland."
 RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF233469; AAF64144.1; -
 DR InterPro: IPR005135; Exo_endo_phos.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 SQ SEQUENCE 362 AA; 40956 MW; D6D607DE521C27E4 CRC64;

Query Match 99.7%; Score 1897; DB 4; Length 362;
 Best Local Similarity 99.7%; Pred. No. 5.3e-150;
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY * 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60
 Db 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60

QY 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 Db 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 QY 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 Db 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 QY 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 Db 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 QY 241 LKWLKKMQAPESATVIFAGDTNLRDEVTRCGGLPNNIYDVWEFLGKPKHCQYTWTDQ 300
 Db 241 LKWLKKMQAPESATVIFAGDTNLRDEVTRCGGLPNNIYDVWEFLGKPKHCQYTWTDQ 300
 QY 301 MNSNLGITTAACKLRFDRIFFRAAABEGHIIIPRSLDLGLEKLDGGRFPDHWGLCLNDI 360
 Db 301 MNSNLGITTAACKLRFDRIFFRAAABEGHIIIPRSLDLGLEKLDGGRFPDHWGLCLNDI 360
 QY 361 IL 362
 Db 361 IL 362

RESULT 3

Q9N9K5 PRELIMINARY; PRT; 362 AA.
 AC Q9N9K5;
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE CDNA FLJ11306 fis, clone PLACE1010031.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
 RA Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
 RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
 RA Matanabe S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,
 RA Niimiya K., Iwayanagi T.;
 RT "MEDO human cDNA sequencing project."
 RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK002168; BAA92119.1; -
 DR InterPro: IPR005135; Exo_endo_phos.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 SQ SEQUENCE 362 AA; 40865 MW; 8E1702C93BCB62E3 CRC64;

Query Match 99.5%; Score 1893; DB 4; Length 362;
 Best Local Similarity 99.7%; Pred. No. 1.1e-149;
 Matches 361; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60
 Db 1 MELGSCLEGREAAEEGEPEVKRRLLCVFASVASCDAVAACFLAENDMEMERALS 60
 QY 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 Db 61 YFEPVEESALERRPETISEPKTYVDLTNEETDSTTSKISPSEDTQOENGSMFSLITWN 120
 QY 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 Db 121 IDGLDNLNLSERAGVCSYALYSPDVIFLQEVIPPYYSYLKRRSSNYEITIGHEGYFT 180
 QY 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240
 Db 181 AIMLKSRVKLKSQELIIPPESTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAARMNQ 240

QY 241 LKAVLKRMQEPASATVITAGDTNLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTQ 300
DB 241 LKAVLKRMQEPASATVITAGDTNLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTQ 300
QY 301 MNSNLGITTAACKLRFRIRIFPRAAEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 360
DB 301 MNSNLGITTAACKLRFRIRIFPRAAEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 360
QY 361 IL 362
DB 361 IL 362

RESULT 4
Q9JUX7 PRELIMINARY; PRT; 370 AA.
AC Q9JUX7; 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE Putative TRAF and TNF receptor associated protein.
GN TRAF OR TRRAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20309820; PubMed=10764746;
RA Pyre S., Declercq W., Ibrahim N., de Boer M., Vandenabeele P.,
RA Van Riettschoten A.G.I., Dewulf N.,
RA Huybreck D., Remacle J.E.,
RT "TRAP, a novel protein that associates with CD40, Tumor Necrosis
RT factor (TNF) Receptor-75 and TNF Receptor-Associated Factors (TRAFs),
RT and that inhibits Nuclear Factor-kappaB activation.";
RL J. Biol. Chem. 275:18586-18593(2000).
DR EMBL: AJ251328; CAB92971.1; -
DR MGD: MGI:1860486; Ttrap.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
KW Receptor.
SQ SEQUENCE 370 AA; 41033 MW; A773A8889DF5BE83 CRC64;

Query Match 66.1%; Score 1258; DB 11; Length 370;
Best Local Similarity 68.4%; Pred. No. 1e-96;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;

QY 12 EAEEEGEPVKRRRLCYEFASVASCDAVAQCFLAENDMERALNSYFEPVEESAL 71
DB 24 EAAQAE-EDRVKRRRLQCLGFLVGGCDPTWVPSVLRNDWQTKALSAVFELPENDQGW 82
QY 72 ERREPITSEKRYVDLJNETTDTSTSKISPEDTQOENGSMESLTTWIDGLDNLNLS 131
DB 83 PROPTPSFKSEAVVDLJNEDANDLTLEASPS-GTFLDSSSTISFTWIDGLDGCNLF 141
QY 132 RANGVCSYALSPDYVFLQEVLPYRSTLKRSSWEITITGHEGYPFALMKSRVKL 191
DB 142 RANGVCSYALSPDYVFLQEVLPYRSTLKRSSWEITITGHEGYPFALMKSRVKL 201
QY 192 KSOEIIIPFPTKMRNLGVHVNVSNGNELCLMTSHLESTRGHAERMNOLKVLKMOEA 251
DB 202 KSOEIIIPFPTKMRNLGVHVNVSNGNELCLMTSHLESTRGHAERMNOLKVLKMOEA 261
QY 252 PESATVIFAGDTNLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTOMNSNLGITAAC 311
DB 262 PDSTVIFAGDTNLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTOMNSNLGITAAC 321
QY 312 KLRFDRIFFRAAEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 362
DB 322 KLRFDRIFFRAAEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 370

RESULT 5
Q9D7N0 PRELIMINARY; PRT; 157 AA.
AC Q9D7N0; 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DE Tnf and Tnf receptor associated protein.
GN TRRAP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=TONGUE;
MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishi Y.,
RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I.,
RA Saito T., Okazaki Y., Gojohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staudil F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
RA Yashizaki Y.;
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL: AK009089; BAB26063.1; -
DR MGD: MGI:1860486; Ttrap.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 157 AA; 17876 MW; 5920850802FAAE84 CRC64;

Query Match 35.5%; Score 675; DB 11; Length 157;
Best Local Similarity 77.4%; Pred. No. 1.2e-48;
Matches 123; Conservative 18; Mismatches 16; Indels 2; Gaps 1;

QY 204 MMRNLGVHVNVSNGNELCLMTSHLESTRGHAERMNOLKVLKMOEPASATVIFAGDT 263
DB 1 MMRNLGVHVNVSNGNELCLMTSHLESTRGHAERMNOLKVLKMOEPASATVIFAGDT 60
QY 264 NLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTOMNSNLGITTAACKLRFRIFPRAA 323
DB 61 NLDRREYTRCGGLPNNIVDWVEFLGPKHCQYTWDTOMNSNLGITTAACKLRFRIFPRAA 323
QY 324 AEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 362
DB 119 AEEGHIIIPRSLDLGLGKLDGCGRPSDHWGLLCNLDI 370
RESULT 6
Q9XWG3 PRELIMINARY; PRT; 362 AA.
AC Q9XWG3; 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DE Y63D3A.4 protein.
GN Y63D3A.4.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;

```

RN [11]
RP SEQUENCE FROM N.A.
RA White S.;
RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
RN [12]
RP SEQUENCE FROM N.A.
RA MEDLINE=99069613; PubMed=9851916;
RX none;
RT "genome sequence of the nematode C. elegans: A platform for
investigating biology.";
RL Science 282:2012-2018(1998).
DR EMBL: AL032652; CA21707.1; -.
DR InterPro: IPR005135; Exo_endo_phos.
DR Pfam: PF03372; Exo_endo_phos; 1.
SQ SEQUENCE 362 AA; 40893 MW; 7895C4750486652 CRC64;

Query Match 22.1%; Score 420; DB 5; Length 362;
Best Local Similarity 29.3%; Pred. No. 7, 1e-27;
Matches 108; Conservative 75; Mismatches 142; Indels 44; Gaps 10;

OY 7 LEGREAEAE-----GEPEVKRRLCY-----EFASVASCDAVAQCFLAEN 50
DB 11 IEARQMSQDSDEVELEIIDEPEQGLKNSMSDEQKLHEFALITATDEAFQSTLDV 70
OY 51 DWEMERALNSYFEPPVESALERRPETISEPKTYVDLTNEETDSTTSKISPEDTQOEN 110
DB 71 DWDLKALDVFY-----GSEAFARSAAGAAVAGASSMASSGAAMTADLK--- 116
OY 111 GSMFLITWNIDGLDNLNLSERAGVSYLALSPDVFLOEVIIPPYSYLKRSSNYEI 170
DB 117 GFEEVSVMNWDIGDGRSLTLRMAKAVHAKVNNPDILFLQEVVDRLAPIDKLSLYK 176
OY 171 ITGHEE-GYFTAILKRSVKLSOEIIPPESTKMMNLLCVHVSNGNELCMTSHLES 229
DB 177 YYSKGGQYTAIVLSK-MEDVEKHDIYHQNQGMRTLDIGSGIGLKVFLNTHLES 235
OY 230 TRGHAEERNQKLKVKLMQOE---APESATVIFAGDTNLDRREYTRCGLPNNIVDWER 286
DB 236 TREHRPQRCAGQFGCMKQVREIITAQNPGALVFFGDLNLRDEYSR---VPDQVKAWEA 292
OY 287 LGRKHQGYTWDPQMSNLSITAACTLRPRIFFRAAEGHIIIPSLDLGLGLEK-DCG 345
DB 293 AGSDNKTFTWDTFKNDKQGFHGAKKRFDRLYW----SGPLDKVYFTLEGQRIRSC 347
OY 346 RFPSDHML 354
DB 348 CFPDHWAI 356

RESULT 7
O9SA95
ID O9SA95 PRELIMINARY; PRT; 441 AA.
AC O9SA95;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE F25C20.3 protein.
GN F25C20.3.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Vysotskaya V.S., Schwartz J.R., Yu G., Toriumi M., Lenz C., Liu S.,
RA Li J., Lee J.M., Kremetskaia I., Lutos J., Ngan I., Liu A.,
RA Gonzalez A., Altafi H., Araujo R., Chao Q., Conn L., Conway A.B.,
RA Dunn P., Hansen N., Huizar L., Kim C., Palm C., Rowley D., Shinn P.,
RA Walker M., Davis R.W., Ecker J.R., Federici N.A., Theologis A.;
RA "Arabidopsis thaliana chromosome 1 BAC F25C20 sequence.";
RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.

```

```

RN [12]
RP SEQUENCE FROM N.A.
RA STRAIN=CV. COLUMBIA;
RA Theologis A.;
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC007296; AAD30241.1; -.
DR InterPro: IPR001969; Asparticase-site.
DR InterPro: IPR005135; Exo_endo_phos.
DR InterPro: IPR001876; znf_RangDP.
DR Pfam: PF03372; Exo_endo_phos; 1.
DR SMART: SM00547; znf_RBZ; 2.
DR PROSITE: PS00141; ASP_PROTEASE; UNKNOWN_1.
DR PROSITE: PS01358; ZF_RANBP2_1; UNKNOWN_1.
SQ SEQUENCE 441 AA; 48995 MW; 6582FAA2D7BD0CA8 CRC64;

Query Match 12.1%; Score 231; DB 10; Length 441;
Best Local Similarity 30.2%; Pred. No. 5, 3e-11;
Matches 84; Conservative 39; Mismatches 109; Indels 46; Gaps 11;

OY 113 MFSLITWNIDGLDNLNLSERAGVSYLALSPDVFLOEVIIPPYSYLKR-----SS 166
DB 175 ILSYVWRREDLELN---LNRRAIGHLIQLSPHLICQEVTPPELYDIFKSNMKAYSC 231
OY 167 NYEITGHEGYFTAIMLKSRSVKLSOEIIPPESTKMMNLLCVHVSNGNE-LCLMTS 225
DB 232 SVSDVAASRGYCYMLL---SKLGVSFSKSFSGNSINGRELSIAEVEVPGKPLVFATS 288
OY 226 HLES-----TRGHAEERNQKLKVKLMQOEAPESATVIFAGDTNLDRREYTRCG-- 275
DB 289 HLESPPGPPKMDQFSREVRQAEAEIILR---PNANVIFGGMMWCD---KLQK 341
OY 276 -LPNNIVDWERFLGPKKQGYTWDPQMSNLSITAACTLRPRIFFR--- 321
DB 342 PLPDKNVDWEVL-KPGDLGFTYDTRKANPMLSGNALOKRDLRILCRDVKLGIEWVG 400
OY 322 AAEGHIIIPSLDLG-LEKIDGRFSDHMLCNL 358
DB 401 KEAIPGLSYKKEKRYGDIKLELPVLSHDHGLLVLT 438

RESULT 8
O9VGS4
ID O9VGS4 PRELIMINARY; PRT; 449 AA.
AC O9VGS4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, last annotation update)
DE CG4796 protein.
GN CG4796.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=BERKELEY;
RA MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.C., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abri1 J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,
RA Butlis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferliera S., Fleischmann W.,

```


Query Match	5.7%;	Score 108;	DB 16;	Length 253;
Best Local Similarity	19.5%;	Pred. No. 0.44;		
Matches	56;	Conservative	52;	Mismatches 119;
			Indels	60;
			Gaps	12;

RESULT 14	
Q8UNX2	
ID Q8UNX2	PRELIMINARY;
AC Q8UNX2;	PRT; 863 AA

Query Match	5.78;	Score 107.5;	DB 15;	Length 863;
Best Local Similarity	21.28;	Pred. No. 2.7;		
Matches	66;	Conservative	40;	Mismatches 107; Indels 99; Gaps 13;

RESULT 15	
Q9CMM2	
ID Q9CMM2	PRELIMINARY; PRT; 267 AA

DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Hypothetical protein PM0691.
GN PM0691.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxId=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.-J., Zhang Q., Li L.-L., Paustian M.-L., Whittam T.S., Kapur V.

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:40 ; Search time 7.56388 Seconds
(without alignments)
2028.884 Million cell updates/sec

Title: US-09-697-863A-4

Perfect score: 1968
Sequence: 1 MASSSSSDAEPAGPAGRAA.....DCGRFSDHGLCTLVNVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	129.5	6.6	388	1 NOCT_XENLA	P79942 xenopus lae
2	126	6.4	429	1 NOCT_MOUSE	O35710 mus musculu
3	123	6.2	431	1 NOCT_HUMAN	O9UK39 homo sapien
4	105.5	5.4	267	1 EX3_HAEIN	P44318 haemophilu
5	99.5	5.1	556	1 PHU_LEPIN	P17627 leptospira
6	98	5.0	494	1 PR31_YEAST	P49704 saccharomyc
7	98	5.0	1495	1 AZMG_MOUSE	O61838 mus musculu
8	93.5	4.8	1000	1 Y041_SYNY3	O55445 synechocyst
9	93.5	4.8	1246	1 SKIW_HUMAN	O15477 homo sapien
10	93	4.7	316	1 APEL_MOUSE	P28352 mus musculu
11	91	4.6	317	1 APEL_BOVIN	P23196 bos taurus
12	90.5	4.6	865	1 CYAA_PROMI	O59685 proteus mir
13	90	4.6	266	1 YARD_ECOLI	P30865 escherichia
14	89.5	4.5	477	1 ISCL_YEAST	P40015 saccharomyc
15	89.5	4.5	663	1 VTER_HSVJ	P52462 human herpe
16	89	4.5	317	1 APEL_HUMAN	P27695 homo sapien
17	89	4.5	962	1 UBP4_MOUSE	P35123 mus musculu
18	88.5	4.5	1196	1 XPG_XENLA	P14629 xenopus lae
19	88	4.5	361	1 APPA_DICDI	P51173 dictyosteli
20	88	4.5	362	1 MPA3_HUMAN	P53082 homo sapien
21	88	4.5	549	1 G6P1_HAEIN	P44312 haemophilus
22	87.5	4.4	354	1 ANGE_DROME	O24239 drosophila
23	87.5	4.4	422	1 NSMA_RAT	O9et64 rattus norv
24	87	4.4	342	1 OUEA_LISMO	O8y629 rattus norv
25	87	4.4	875	1 SYA_CHIMU	O9p1h5 chlamydia m
26	86.5	4.4	1503	1 TRU2_HUMAN	O94759 homo sapien
27	85.5	4.3	4344	1 DYHC_EMENT	P45444 emeriella
28	85	4.3	316	1 APEL_RAT	P43138 rattus norv
29	85	4.3	342	1 OUEA_LISTIN	O92b13 listeria in
30	85	4.3	1135	1 RBL2_RAT	O55081 rattus norv
31	85	4.3	1391	1 RPOB_MYCPN	P78013 mycoplasma
32	84.5	4.3	476	1 PPRB_PSEAE	P35483 pseudomonas
33	84.5	4.3	550	1 G6P1_VIBCH	O9kuy4 vibrio chol

34	84.5	4.3	617	1 HEMA_MEASA	P35971 measles vir
35	84.5	4.3	2468	1 MAPB_HUMAN	P46821 homo sapien
36	84	4.3	353	1 ALCI_GORCO	P20758 gorilla gor
37	84	4.3	412	1 ACDS_RAT	P15651 rattus norv
38	84	4.3	573	1 C114_MOUSE	P19467 mus musculu
39	84	4.3	963	1 UBP4_HUMAN	O13107 mus sapien
40	84	4.3	1152	1 MAP4_HUMAN	P27816 homo sapien
41	83.5	4.2	536	1 ARP_ARATH	P45951 arabidopsis
42	83.5	4.2	978	1 PEX6_RAT	P54777 rattus norv
43	83.5	4.2	1017	1 LRGI1_YEAST	P35688 saccharomyc
44	83.5	4.2	2700	1 ZAN_HUMAN	O9y493 homo sapien
45	83	4.2	353	1 DRN2_MOUSE	P56442 mus musculu

ALIGNMENTS

RESULT 1	ID	NOCT_XENLA	STANDARD:	PRT:	388 AA.
AC	P79942:	15-JUL-1999 (rel. 38, last sequence update)			
DT	15-JUL-1999 (rel. 38, last sequence update)				
DT	15-JUN-2002 (rel. 41, last annotation update)				
DE	Nocturnin (Rhythmic message 1) (RMI).				
OS	Xenopus laevis (African clawed frog).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;				
OC	Xenopodinae; Xenopus.				
OX	NCBI_TaxID=8355;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Retinal photoreceptors;				
FX	MEDLINE=97121484; PubMed=8962150;				
RA	Green C.B., Besharse J.C.;				
RT	"Identification of a novel vertebrate circadian clock-regulated gene encoding the protein nocturnin."				
RL	Proc. Natl. Acad. Sci. U.S.A. 93:14884-14888(1996).				
CC	- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR				
CC	OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH				
CC	MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT				
CC	LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES.				
CC	- TISSUE SPECIFICITY: EXPRESSED ONLY IN THE PHOTORECEPTORS OF THE				
CC	RETINA. EXPRESSION IS CONTROLLED BY THE RETINAL CIRCADIAN CLOCK.				
CC	- SIMILARITY: BELONGS TO THE CCR4/NOCTURNIN FAMILY.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/				
CC	or send an email to license@sib-sib.ch).				
CC	-----				
DR	EMBL: U74761; AAB39495.1; "				
DR	InterPro: IPR005135; Exo_endo_phos.				
DR	Pfam: PF03372; Exo_endo_phos; 1.				
KW	Biological rhythms.				
SO	SEQUENCE 388 AA; 43940 MW; 7c9f46f28d9c4e7e CRC64;				
Query Match					
Best local similarity 6.6%; Score 129.5; DB 1; Length 388;					
Matches 72; Conservative 50; Mismatches 158; Indels 57; Gaps 14;					
OY	74	ELPENNQGWPRPPTFSKSAVYVDLNFNEA---NPTTLEASPSTPLDSDTISFTWN 130			
DB	51	EASGHDSFPLD-PKELDLBCOVALODRPARLRHDFSLRSSSS---QQRPTFRWQMN 106			
OY	131	I-----DGLDG---CNL-----PERARGVSCALYSFVVFLOEV----- 163			
DB	107	ILAQALGEGCDNIMCPMALKWEKRYDLEELIMYQPDVLCLOEVDHYHPTQPILSR 166			
OY	164	IPPYCAVILKKRASVITITIGNEGYFTAILLKKGRVKKFSQRIIPPTMKMRNLGVNV 223			

Db	167	LGYYCTTLAKWMSCLDVEYHNHNGDGCALFELDORFQLVNSAKIRLSAPRTKTNQVAIA	2266
Oy	224	SL-----GGEFCLMTSHLESTREHSABRIQOLKTVLCKMQEADPTSTV--IFAGDTNLRD	2777
Db	227	TLOCCETGRQLCFVAVTHLKAKTGWERRRLAGSDLLDNLESITOGATVAPLIGDFNADP	2866
Oy	278	QEVLT--KCGGLPDVVPAMEFL--GKPKHQOYWTDKRANNLNRIPIAYKHREDRIEFA	3322
Db	287	TEEYVKRRASSLMLNSAYKLLSDGSEBPPYTTW-----KIRTTGESCHLIDYIWY-S	3393
Oy	333	EEGHILPOSIDL-----VGLKLDGCRFPSPDHMLLC	364
Db	340	QHALRVNAALIGLPTEEQIGPKNRLSEFNPSHLSLVC	376

RESULT 2

ID	NOCT_MOUSE	STANDARD	PRJ	429 AA.
AC	035710; Q8QZ69;			
DT	15-JUL-1999 (Rel. 38, Created)			
DT	15-JUN-2002 (Rel. 41, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Nocturnin (CCR4 protein homolog).			
CN	CCRN4L OR NOC OR CCR4.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxId=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/c; TISSUE-Brain;			
RX	MEDLINE-99453012; PubMed-10521507;			
RA	Dupressoir A., Barbot W., Loireau M.P., Heidmann T.;			
RT	"Characterization of a mammalian gene related to the yeast CCR4			
RT	general transcription factor and revealed by transposon insertion.";			
RL	J. Biol. Chem. 274:31068-31075(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-BALB/c; TISSUE-Retina;			
RA	Wang Y., Osterbur D.L., Green C.B., Besharse J.C.;			
RT	"Mammalian homologs of Xenopus nocturnin: conservation of structure			
RT	and circadian regulation.";			
RL	Submitted (OCR-1999) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE OF 62-429 FROM N.A.			
RC	STRAIN-DBA/2J; TISSUE-Liver;			
RX	MEDLINE-97190339; PubMed-9038221;			
RA	Puech A., Dupressoir A., Loireau M.P., Mattei M.-G., Heidmann T.;			
RT	"Characterization of two age-induced intracellular A-particle-related			
RT	transcripts in the mouse liver. Transcriptional read-through into an			
RT	open reading frame with similarities to the yeast ccr4 transcription			
RT	factor.";			
RL	J. Biol. Chem. 272:5995-6003(1997).			
CC	-1- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR			
CC	OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH			
CC	MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT			
CC	LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES (BY SIMILARITY).			
CC	-1- SIMILARITY: BELONGS TO THE CCR4/NOCTURNIN FAMILY.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sdb.ch/announce/			
CC	or send an email to license@sdb-sdb.ch).			
CC	-----			
DR	EMBL; AF183960; AAG56547.1; -			
DR	EMBL; AF198491; AAG01384.1; -			
DR	EMBL; U70139; AAB62717.1; ALT_FRAME.			
DR	MGI; MGI:109382; Ccr4.			
DR	InterPro: IPR005135; Exo_endo_phos.			

DR	Pfam: PF03372; Exo-endo_phos. 2.		
KW	Biological rhythms.		
FT	CONFLICT 123	YOR -> LPA (IN REF. 3.)	
SO	SEQUENCE 429 AA; 48300 MW; CE9FE5DD4E13943 CRC64;		
Query Match	6.4%;	Score 126;	DB 1; Length 429;
Best Local Similarity	21.8%;	Pred. No. 0.0019;	
Matches 88;	Conservative 55;	Mismatches 180;	Indels 72; Gaps 18;

[illegible]

RESULT 3

CC ID NOCT_HUMAN STANDARD: PRT: 431 AA.
 CC Q9UK39; Q9HD93; Q9HD94; Q9HD95;
 CC DT 15-JUN-2002 (Rel. 41, Created)
 CC DT 15-JUN-2002 (Rel. 41, Last sequence update)
 CC DT 15-JUN-2002 (Rel. 41, Last annotation update)
 CC DE Nocturnin (CCR4 protein homolog).
 CC CCR4L OR NOC OR CCR4.
 CC OS Homo sapiens (Human).
 CC OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Primates; Catarrhini; Hominiade; Homo.
 CC OX NCBI_TaxID=9606;
 CC RN [1]
 CC RN SEQUENCE FROM N.A.
 CC RX MEDLINE=99453012; PubMed=10521507;
 CC RA Dupressoir A., Barbot W., Loiseau M.P., Heidmann T.;
 CC RT "Characterization of a mammalian gene related to the Yeast CCR4
 CC J. Biol. Chem. 274:31068-31073(1999).
 CC RN [2]
 CC RP SEQUENCE OF 67-431 FROM N.A.
 CC RA Wang Y., Osterbur D.L., Green C.B., Besharse J.C.;
 CC RT "Mammalian homologs of Xenopus nocturnin: conservation of structure
 CC and circadian regulation.";
 CC RL Submitted (Oct-1999) to the EMBL/Genbank/DBJ databases.
 CC -1- FUNCTION: COMPONENT OF THE CIRCADIAN CLOCK OR DOWNSTREAM EFFECTOR
 CC OF CLOCK FUNCTION. EXHIBITS A HIGH AMPLITUDE CIRCADIAN RHYTHM WITH
 CC MAXIMAL LEVELS IN EARLY EVENING. IN CONSTANT DARKNESS OR CONSTANT
 CC LIGHT, THE AMPLITUDE OF THE RHYTHM DECREASES (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CCR4/NOCTURNIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
 CC the European Bioinformatics Institute. There are no restrictions on its

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF183961; AAD56548.1; -
DR EMBL; AF199492; AAG01387.1; -
DR EMBL; AF199493; AAG01388.1; -
DR EMBL; AF199494; AAG01389.1; -
DR Genbank; HGNC:14254; CCRN4L.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 3.
KW Biological rhythms.
FT CONFLICT 69 69 T -> N (IN REF. 2).
FT CONFLICT 77 77 G -> A (IN REF. 2).
FT CONFLICT 266 266 A -> T (IN REF. 2; AAG01389).
FT CONFLICT 341 341 N -> S (IN REF. 2).
SQ SEQUENCE 431 AA; 48150 MW; B61EF484E8D29AF5 CRC64;

Query Match 6.28; Score 123; DB 1; Length 431;
Best Local Similarity 21.58; Pred. No. 0.0036;
Matches 90; Conservative 56; Mismatches 181; Indels 92; Gaps 20;

QY 6 SSDAEPAGPAGRAASAPAEAAQAEEDRVKRRRLQCLGFA--LVGGCDPTWVPSVLEND 62
DB 33 SPRAVPRPAPRLLAASASGAA-RSCSTVCSMGIGTSRLTSGLAKTINSSAASOH- 90
QY 63 WOTQALSAVYELPENNCGWPRQP-----PTSFKEAYVDLTNEDANDTTI 108
DB 91 -----PEYIVSPDEHLEPIDPELLEECRAVLHTRPPRFQND-FVDLMTD----- 135
QY 109 LEASPSGFLDSSSTISFTWNI-----DGLDG--CNLPERARGVSCIAL-----YS 154
DB 136 ---CPSTHP-----PIRWQNNMLIAQALGEEKDNFVOCPEVALKWEKKCILLLEIAYQ 187
QY 155 PDVYFLQEV-----IPPYCA-----YLKRAASYTIITGNEGYFTATLLKGVKF 201
DB 188 PDILCLQGVHDYFDTFQFLSLRLYQGFPPKWPSPCLDVENHNGPPOCALFVLQNFKL 247
QY 202 KSOELIPEPPNKKMARNLCLVNSL---GNEPCLMTSHLESTREHSAERIRQLKTVIGK 257
DB 248 VNSANIRLTAMTLKTNOVAIAQTECKESGRQFCIAVTHLKARFGWERFRSAGCDDLQN 307
QY 258 MQEAPDSTTV--IFAGPTNLR-DOEVIKCGGLPD-NVFDAMEFL--GKRHCQYTWDTK 310
DB 308 LQNTQGGKIRILYCGDFNAEPTEVYKHHASSMNLNLSADQSEPPYTTWKIR 367
QY 311 ANNMLRIPAAVYKRRFDRIFFRAEBGHLIPQSLD---VGLKLDGCRFPSPDHGGLIC 364
DB 368 TSGECR-----HTLDYIYW-SKHALWRSALDLTTEQIGPNRLPSFNYPSSHLSLVC 419

RESULT 4
EX3_HAEIN STANDARD; PRT; 267 AA.
AC P44318;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Exodeoxyribonuclease III (EC 3.1.11.2) (Exonuclease III) (EXO III).
GN XTHA OR H10041.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
ON NCBI_TaxID:727;
RX MEDLINE=95350630; Pubmed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

```

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Ulfersack T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Eriichman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -1- FUNCTION: MAJOR APURINIC-APYRIMIDINIC ENDONUCLEASE OF E.COLI. IT
CC REMOVES THE DAMAGED DNA AT CYTOSINES AND GUANINES BY CLEAVING ON
CC THE 3' SIDE OF THE AP SITE BY A BETA-ELIMINATION REACTION
CC (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Degradation of double-stranded DNA. It acts
CC progressively in a 3'-to 5'-direction, releasing nucleoside 5'-
CC phosphates.
CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32689; AAC21719.1; -
DR HSSP; P09030; IAKO.
DR TIGR; HT0041; -.
DR InterPro; IPR000097; Apendonclisel.
DR InterPro; IPR004442; ExonNase_III.
DR InterPro; IPR004808; ExoIII_xth.
DR InterPro; IPR005135; Exo_endo_phos.
DR Pfam; PF03372; Exo_endo_phos; 1.
DR TIGRFRAMS; TIGR00195; exoNase_III; 1.
DR TIGRFRAMS; TIGR00633; xth; 1.
DR PROSITE; PS00726; AP_NUCLEASE_F1_1; 1.
DR PROSITE; PS00727; AP_NUCLEASE_F1_2; 1.
DR PROSITE; PS00728; AP_NUCLEASE_F1_3; 1.
DR Hydrolase; Nuclease; Exonuclease; DNA repair; Complete proteome.
FT METAL 34 34 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT ACT_SITE 259 259 GENERAL BASE (BY SIMILARITY).
SQ SEQUENCE 267 AA; 31025 MW; 6B3ADE465A1E347C CRC64;

Query Match 5.4%; Score 105.5; DB 1; Length 267;
Best Local Similarity 22.2%; Pred. No. 0.062;
Matches 63; Conservative 43; Mismatches 105; Indels 73; Gaps 14;

QY 124 ISFTTNIDGIDGCLNPERARGVSCALYSPDVVFLQEV-VIPPYCAVLLKRAASYTI 181
DB 1 MKFISFNINGLRA---RPHOLEAITEIKYQDPYDGLQEIKAIVADBAFYEITENGLYHF 55
QY 182 TGNEEGYFTAILLKGRVAFKSGOEIIPPTNKKMARNLCLVNSLGNFCLMT----- 234
DB 56 HIGQKHGYVALLTKOEPV-IRGEPPTNEDAQRILIMADE---TEGGLTVINGYRP 111
QY 235 ---SHLESTREHSAERI-RQLTVLQKMQEAPDSTTVIFAGPTNRDQEV----- 281
DB 112 QGSRHAHEKFPFAKEFVADQYLEK--EHDKSNPILMGDMNISPDLDIGIDENRK 169
QY 282 -----KCGGLPD-----NVFDAMEFLGKPKHCQYTW-DTKA-----NNMLRIP 318
DB 170 RWLRGCKCSFLPERAWYORLYDGLSDSFRKLIPTANDKFSWFDYRSKGDNDGLNI- 228
QY 319 AAYKIHFRDIFPR---AEEGHLIPQSLDVGLEKLDGCRFPSDH 359
DB 229 -----DHILVSQKLAERCVDVGIALDRAMEK-----PSDH 259

RESULT 5
PHL_LEPIN STANDARD; PRT; 556 AA.
AC P17627;

```


RESULT 7
 A2MG_MOUSE STANDARD; PRT; 1495 AA.
 AC 061838; 060628;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 01-NOV-1997 (Rel. 35, Last annotation update)
 DE Alpha-2-macroglobulin precursor (Alpha-2-M).
 GN A2M.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-44 AND 1240-1259.
 RC TISSUE=Liver;
 RX MEDLINE=93076803; PubMed=1280217;
 RA van Leuven F., Torrekens S., Overbergh L., Lorent K., de Strooper B.,
 van den Berghe H.;
 RT "The primary sequence and the subunit structure of mouse alpha-2-
 macroglobulin, deduced from protein sequencing of the isolated
 subunits and from molecular cloning of the cDNA."
 RL Eur. J. Biochem. 210:319-327(1992).
 RN [2]
 RP SEQUENCE OF 1-161 FROM N.A.
 RC STRAIN=129/J;
 RX MEDLINE=95095249; PubMed=7528166;
 RA Umans L., Sernaeels L., Hilliker C., Stas L., Overbergh L.,
 de Strooper B., van Leuven F., van den Berghe H.;
 RT "Molecular cloning of the mouse gene coding for alpha 2-macroglobulin
 and targeting of the gene in embryonic stem cells."
 RL Genomics 22:519-529(1994).
 CC -I- FUNCTION: IS ABLE TO INHIBIT ALL FOUR CLASSES OF PROTEINASES BY A
 UNIONE "TRAPPING" MECHANISM. THIS PROTEIN HAS A PEPTIDE STRETCH,
 CALLED THE "BAIT REGION", WHICH CONTAINS SPECIFIC CLEAVAGE SITES
 FOR DIFFERENT PROTEINASES. WHEN A PROTEINASE CLEAVES THE BAIT
 REGION, A CONFORMATIONAL CHANGE IS INDUCED IN THE PROTEIN WHICH
 TRAPS THE PROTEINASE. THE ENTRAPPED ENZYME REMAINS ACTIVE AGAINST
 LOW MOLECULAR WEIGHT SUBSTRATES (ACTIVITY AGAINST HIGH MOLECULAR
 WEIGHT SUBSTRATES IS GREATLY REDUCED). FOLLOWING CLEAVAGE IN THE
 BAIT REGION A THIOLESTER BOND IS HYDROLYZED AND MEDIATES THE
 COVALENT BINDING OF THE PROTEIN TO THE PROTEINASE.
 CC -I- TISSUE SPECIFICITY: PLASMA.
 CC -I- DEVELOPMENTAL STAGE: CONTRARY TO THE RAT PROTEIN, WHICH IS AN
 ACUTE PHASE PROTEIN, THIS PROTEIN IS ALWAYS PRESENT AT HIGH LEVELS
 IN CIRCULATION.
 CC -I- SIMILARITY: TO OTHER PROTEINS OF THE ALPHA-MACROGLOBULIN FAMILY,
 INCLUDING COMPLEMENT COMPONENTS C3, C4, AND C5.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M93264; AAA39508.1;
 CC EMBL: U06977; AAA87890.1;
 CC HSSP: F01023; 1BVB.
 CC MGD: MGI:87854; A2M.
 CC DR InterPro: IPR002890; A2M_N.
 CC DR InterPro: IPR001599; Macroglb1na2.
 CC DR Pfam: PF00207; A2M; 1.
 CC DR Pfam: PF01835; A2M_N; 1.
 CC DR PROSITE: PS00477; ALPHA_2_MACROGLOBULIN; 1.
 CC KW Serine protease inhibitor; Glycoprotein; Plasma; Bait region; Signal.
 CC SIGNAL 1 24
 CC CHAIN 25 1239 POTENTIAL.
 CC CHAIN 1240 1495 ALPHA-2-MACROGLOBULIN 165 KDA SUBUNIT.
 CC DOMAIN 686 744 BAIT REGION.

FT DISULFID 48 86 BY SIMILARITY.
 FT DISULFID 249 298 BY SIMILARITY.
 FT DISULFID 267 286 BY SIMILARITY.
 FT DISULFID 277 430 BY SIMILARITY.
 FT DISULFID 594 783 BY SIMILARITY.
 FT DISULFID 642 689 BY SIMILARITY.
 FT DISULFID 833 861 BY SIMILARITY.
 FT DISULFID 859 895 BY SIMILARITY.
 FT DISULFID 933 1339 BY SIMILARITY.
 FT DISULFID 1092 1140 BY SIMILARITY.
 FT THIOLEST 984 987 BY SIMILARITY.
 FT CARBOHYD 55 55 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 157 157 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 405 405 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 412 412 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 881 881 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 942 942 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1003 1003 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1385 1385 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1443 1443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CONFLICT 41 47 OESLRP -> SGIPERA (IN REF. 2).
 FT CONFLICT 91 91 I -> V (IN REF. 2).
 FT CONFLICT 142 144 VKF -> GIT (IN REF. 2).
 FT CONFLICT 160 161 FP -> VS (IN REF. 2).
 SO SEQUENCE 1495 AA; 165827 MW; 0D21568B8B0B4CCB CRC64;
 Query Match 5.0%; Score 98; DB 1; Length 1495;
 Best Local Similarity 22.9%; Pred. No. 3.3; Indels 82; Gaps 15;
 Matches 66; Conservative 34; Mismatches 106;
 QY 26 AQAEEDRYKRRRLQGLFALVG-----CDPTWPSVLRND--WQO-----KALS 70
 DB 1149 SQOSESHVYTKALLAYAFALAGNKKARSELLESINKAVKEEDSLHWQPGDVQVKALS 1208
 QY 71 AYFELPNDQMPQPPPTSESEAYV---DLTNDADDTTLTASPGTLEDSSTIS-F 126
 DB 1209 FY-----QPRPSAEVMTAYVLLAYLTSSSRPTDLSS-----DLSTASKI 1252
 QY 127 ITW-----NIDLDGCMLEPARAGVCSCLALYSPDVVFLQEVIPYCAVLLKRAASTII 181
 DB 1253 VKWISKOONSNG-----GL-----LLPDYVALQALSKYSASFPTSQKREVLV 1296
 QY 182 TGENRGYFTAILKKGRYKESQEI---IP-----FPNTKMRRLLCVNS 224
 DB 1297 TSSSGTFSTKPFHNSGNRLLEQVRLPDLGNYVTGSGGCVYLTQSLKNYL--PVA 1354
 QY 225 LGNMEFCMTSHL-----ESTREHSAERTRQLKTYLGMQKEAPDSTTVI 268
 DB 1355 DGKAPFALQVNTLPNFDKAEHDRTFOIRINWSTYG---ERPSSNMVI 1399
 RESULT 8
 Y041_SYNY3 STANDARD; PRT; 1000 AA.
 AC Q55445;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DE Hypothetical methyl-accepting chemotaxis protein sll0041.
 GN SLL0041.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
 OX NCBI_TaxID=1148;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96127529; PubMed=8590279;
 RA Kaneko T., Tanaka A., Sato S., Kotani H., Sazuka T., Miyajima N.,
 Sugita M., Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. I. Sequence features in the 1 Mb
 region from map positions 64k to 92k of the genome."
 RT

```

RL DNA Res. 2:153-166(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-97061201; PubMed-8905231;
RA Kaneo T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirasawa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSOR TRANSDCERS FAMILY.
CC -1- SIMILARITY: BELONGS TO THE PHYTOCHROME FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D64006; BAA10787.2; -.
DR HSSP; P02942; 1OU7.
DR InterPro: IPR004089; Chntaxis_transd.
DR InterPro: IPR003018; GAF.
DR InterPro: IPR003660; HAMP.
DR InterPro: IPR004090; Me_chemotaxis.
DR InterPro: IPR001294; Phytochrome.
DR Pfam; PF00015; MCPsignal; 1.
DR Pfam; PF00672; HAMP; 1.
DR Pfam; PF01590; GAF; 2.
DR PRINTS; PR00260; CHEMTRNSDCR.
DR SMART; SM00065; GAF; 2.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
DR PROSITE; PS50046; PHYTOCHROME.2; 2.
DR Hypothetical protein; Transducer; Complete proteome.
SQ SEQUENCE 1000 AA; 108334 MW; 58C5A765E1275FDF CRC64;
KW
Query Match 4.8%; Score 93.5; DB 1; Length 1000;
Best Local Similarity 19.8%; Pred. No. 4.7;
Matches 78; Conservative 56; Mismatches 121; Indels 139; Gaps 18;
QY 1 MASGSSDAEAPGAPGRASAPPEAAQAEEDVKRRRLQCLGALVGGCDPMVPSVLR 60
DB 17 VANGETLDGALTFTGLTDAADDESSES-----ASFATIDGDKSEVGDAL-- 63
QY 61 NNM-----OTOKALSAVFELPENDGMPRPPTSSEAVY 96
DB 64 -DMFSGKRETDINGEADGADGIADTQAKTLVLENITEPEID-----GAEFMAEAFI 115
QY 97 DLTNEADNTTLEASPSGTPLEDSTTISFTWNIDGL--DGCNLP-----RANGVSC 149
DB 116 -----AEMTAVEDVSPNPNAIDTDAALFQSAVELPPRPINLPKVELPPMQPLAPL 169
QY 150 LALYSPDVVFLDEVIPRYAYLAKRAASTYTTTGNEEGFTAILLK-----GVYKRS 203
DB 170 MALADD-----NLSPMSTSIQAPTOGSGSLRN-KAVLLDLLGLIPAGVIGGLINS 222
QY 204 OETIIPPNTKMMRNLLCVAVSLGNEFCMLTSHLESTRHSERIRQLKTVIGKMQEAPD 263
DB 223 VDLPLPQTE-----QQVKD 237
QY 264 STTVIFAGDTNLRDQEVIKGGLPDNV--PDAMEFLGPKHQYTWPTKANNNIRIPAA 320
DB 238 STT-----KQIRDOILL--GLLVYAVGAFAVAYMNGE-----NTKQATLALAKK 281
QY 321 YKMR-FDR-IFPRAEGHILIPQSLDVG--LEKL 350
DB 282 HSHRNLDPLAVAGDELAIADQITDALSNQVEKL 315

```

```

RESULT 9
SKIW_HUMAN STANDARD; PRT; 1246 AA.
ID SKIW_HUMAN
AC Q15477; Q12902; O15005; Q15476;
DT 01-NOV-1997 (Rel. 35, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Helicase SKI2W (Helicase-like protein) (HLP).
GN SKIV2L OR SKIV2 OR SKI2W OR DDX13 OR W.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-95334363; PubMed-7610041;
RA Dangel A.W., Shen L., Mendoza A.R., Wu L.-C., Yu C.Y.;
RT "Human helicase gene SKI2W in the HLA class III region exhibits
RT striking structural similarities to the yeast antiviral gene SKI2 and
RT to the human gene KIA0052: emergence of a new gene family.";
RL Nucleic Acids Res. 23:2120-2126(1995).
RN [2]
RP SEQUENCE FROM N.A.
RA Bowen L., Dankers C., Baskin D., Faust J., Loretz C., Ahearn M.E.,
RA Banta A., Swartzell S., Smith T.M., Spies T., Hood L.;
RT "Sequence determination of 300 kilobases of the human class III MHC
RT locus.";
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RA Albertella M.R., Jones H., Thomson W., Olavesen M.G.,
RA Campbell R.D.;
RT Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE OF 1-958 FROM N.A.
RA TISSUE-Hippocampus;
RC TISSUE-Hippocampus;
RA Lee S.-G., Song K.;
RT "Cloning of the human helicase-like protein gene homologous to Yeast
RT SKI2.";
RL Korean J. Biochem. 26:215-219(1994).
RN [5]
RP SEQUENCE OF 230-960 FROM N.A.
RA TISSUE-Liver;
RC TISSUE-Liver;
RX MEDLINE-95278935; PubMed-7759100;
RA Lee S.-G., Lee I., Park S.H., Kang C., Song K.;
RT "Identification and characterization of a human cDNA homologous to
RT yeast SKI2.";
RL Genomics 25:660-666(1995).
RN [6]
RP REVISIONS.
RA Song K.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [7]
RP FUNCTION: HELICASE; HAS ATPASE ACTIVITY.
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: BELONGS TO THE HELICASE FAMILY. SKI2 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z48796; CAAB8733.1; ALT_INIT.
DR EMBL; AF019413; AAB67978.1; -.
DR EMBL; X98378; CAAB7024.1; -.
DR EMBL; U09877; AAB52523.1; ALT_INIT.
DR Genew; HGNC:10998; SKIV2L.
DR MIM; 600478; -.
DR InterPro; IPR001410; DEAD.

```

InterPro: IPR001650; Helicase_C.
 DR Pfam: PF00270; DEAD; 1. Helicase_C; 1.
 DR SMART: SM00487; DEXDC; 1.
 DR SMART: SM00490; HELIC; 1.
 KW Helicase; ATP-binding; RNA-binding.
 FT NP_BIND 332 339
 FT SITE 423 426
 FT CONFLICT 66 120
 FT CONFLICT 151 151
 FT CONFLICT 253 253
 FT CONFLICT 265 265
 FT CONFLICT 289 289
 FT CONFLICT 366 366
 FT CONFLICT 562 583
 FT CONFLICT 623 623
 FT CONFLICT 914 914
 FT CONFLICT 917 917
 FT CONFLICT 947 960
 FT CONFLICT 1052 1052
 FT CONFLICT 1071 1071
 SQ SEQUENCE 1246 AA; 137799 MW; 904405C8C70621A0 CRC64;
 Query Match 4.8%; Score 93.5; DB 1; Length 1246;
 Best Local Similarity 21.2%; Pred. No. 6.4;
 Matches 74; Conservative 36; Mismatches 108; Indels 131; Gaps 17;
 QY 7 SDAAEAGPAGRAASAPLAAQAEEDVKKRRRLQCLGALV---GGCDPTMV----- 54
 DB 882 SDDPDRGRPTAEVPPD-----DLVGEFKFLPEGGPDHTVVKLQPGDMA 927
 QY 55 --PSVLRNDWOTOKALSAVFELPNDQGMPPPTSFSEAYVDITNEDANDITILEAS 112
 DB 928 ITTKVLAVNG---EKLLEDFSK-----KQQPKFKDPLAVTAVOELLRLAQM 975
 QY 113 PGGTP-----LEDSSTISFIWNIDGLDGCNLPKARGVCSCLALSPDVVFLQEV 163
 DB 976 PAPPPLDVEVDLQTKMDSVE-----GGLRARKLELILQ--AQC--VHSFR----- 1019
 QY 164 IPYCAVYLKRAASYIITGNFEGYFTALLKKGRKF--KQSEIIPP-----NKKMRN 217
 DB 1020 FP--AOYLKLR-----MQLQKEMERLRLFLSDQSLPEPEYHQRYEVLR 1064
 QY 218 LILCVN-----VSLGNEFCIMTSH----- 236
 DB 1065 LGVVDAGVYKLAGRYACAMSSHELLITFLMDNALSTIRPEIALLSGLVQSGPDAG 1124
 QY 237 --LESTRHSABRIROKTVLQKMEA--PDSTVIIFAGDINLRDOEVI 281
 DB 1125 DQLPNTLKQGIERYAVAKRIGEVYACGLNQTVEEFVGLNFEV 1173
 RESULT 10
 APEL_MOUSE STANDARD; PRT; 316 AA.
 AC P28352;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DE 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DNA (apurinic or apyrimidinic site) lyase (EC 4.2.99.18) (AP
 endonuclease 1) (APEX nuclease) (APEX).
 GN APEX OR APE.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

EX MEDLINE=92041936; PubMed=1939131;
 RA Seki S., Akiyama K., Watanabe S., Hatsushika M., Ikeda S.,
 RA Tsutsui K.,
 RT "cDNA and deduced amino acid sequence of a mouse DNA repair enzyme
 RT (APEX nuclease) with significant homology to Escherichia coli
 RT exonuclease III.";
 RT J. Biol. Chem. 266:20797-20802(1991).
 RN [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=129; TISSUE=Embryo;
 RX MEDLINE=95178846; PubMed=7533013;
 RA Takiguchi Y., Chen D.J.,
 RT "Genomic structure of the mouse apurinic/apyrimidinic endonuclease
 RT gene.";
 RL Mamm. Genome 5:717-722(1994).
 RN [3]
 RN SEQUENCE FROM N.A.
 RC STRAIN=BA16/C; TISSUE=Blood;
 RX MEDLINE=95301294; PubMed=7782087;
 RA Akiyama K., Nagao K., Oshida T., Tsutsui K., Yoshida M.C., Seki S.,
 RT "Cloning, sequence analysis, and chromosomal assignment of the mouse
 RT Apex gene.";
 RL Genomics 26:63-69(1995).
 RN [4]
 RN PARTIAL SEQUENCE OF 1-21, AND CHARACTERIZATION.
 RP TISSUE=Ascites;
 RX MEDLINE=91363416; PubMed=1716153;
 RA Seki S., Ikeda S., Watanabe S., Hatsushika M., Tsutsui K., Akiyama K.,
 RA Zhang B.,
 RT "A mouse DNA repair enzyme (APEX nuclease) having exonuclease and
 RT apurinic/apyrimidinic endonuclease activities: purification and
 RT characterization.";
 RL Biochim. Biophys. Acta 1079:57-64(1991).
 CC -1- FUNCTION: REPAIRS OXIDATIVE DNA DAMAGES IN VITRO. MAY HAVE A ROLE
 CC IN PROTECTION AGAINST CELL LETHALITY AND SUPPRESSION OF MUTATIONS.
 CC REMOVES THE BLOCKING GROUPS FROM THE 3' TERMINI OF THE DNA STRAND
 CC BREAKS GENERATED BY IONIZING RADIATIONS AND BLEOMYCIN.
 CC -1- CATALYTIC ACTIVITY: The C-O-P bond 3' to the apurinic or
 CC apyrimidinic site in DNA is broken by a beta-elimination reaction,
 CC leaving a 3'-terminal unsaturated sugar and a product with a
 CC terminal 5'-phosphate.
 CC -1- SUBUNIT: MONOMER (BY SIMILARITY).
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC -1- SIMILARITY: BELONGS TO THE AP/EXO FAMILY OF DNA REPAIR ENZYMES.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL: D90374; BAA1382.1; -;
 DR EMBL: U12273; AAC13769.1; -;
 DR EMBL: D38077; BAA07270.1; -;
 DR PIR: S17524; S17524.
 DR PIR: A39500; A39500.
 DR HSSP: P27695; 1BIX.
 DR MGD: MGI:88042; Apex.
 DR InterPro: IPR000097; Apendonc1sel.
 DR InterPro: IPR004442; ExoNase-III.
 DR InterPro: IPR004808; ExoIII_xth.
 DR InterPro: IPR005135; Exo_endo_phos.
 DR Pfam: PF03372; Exo_endo_phos; 1.
 DR TIGRFAMs: TIGR00195; exoNase-III; 1.
 DR TIGRFAMs: TIGR00633; xth; 1.
 DR PROSITE: PS00726; AP_NUCLEASE_F1_1; 1.
 DR PROSITE: PS00727; AP_NUCLEASE_F1_2; 1.
 DR PROSITE: PS00728; AP_NUCLEASE_F1_3; 1.
 KW DNA repair; Lyase; Nuclear protein.
 FT INT_MET 0 0
 FT METAL 66 66 MAGNESIUM OR MANGANESE (BY SIMILARITY).

CC	-1- TISSUE SPECIFICITY: THYMUS.
CC <td>-1- INDUCTION: BY SEVERAL DNA DAMAGING AGENTS.</td>	-1- INDUCTION: BY SEVERAL DNA DAMAGING AGENTS.
CC <td>-1- SIMILARITY: BELONGS TO THE AP/EXO/A FAMILY OF DNA REPAIR ENZYMES.</td>	-1- SIMILARITY: BELONGS TO THE AP/EXO/A FAMILY OF DNA REPAIR ENZYMES.
CC <td></td>	
CC <td>This SWISS-PROT entry is copyright. It is produced through a collaboration</td>	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC <td>between the Swiss Institute of Bioinformatics and the EMBL outstation -</td>	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC <td>the European Bioinformatics Institute. There are no restrictions on its</td>	the European Bioinformatics Institute. There are no restrictions on its
CC <td>use by non-profit institutions as long as its content is in no way</td>	use by non-profit institutions as long as its content is in no way
CC <td>modified and this statement is not removed. Usage by and for commercial</td>	modified and this statement is not removed. Usage by and for commercial
CC <td>entities requires a license agreement (See http://www.isb-sib.ch/announce/</td>	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC <td>or send an email to license@lsb-sib.ch).</td>	or send an email to license@lsb-sib.ch).
CC <td></td>	
DR <td>EMBL: X56685; CAA40014.1; .</td>	EMBL: X56685; CAA40014.1; .
DR <td>PIR: S26830; S26830.</td>	PIR: S26830; S26830.
DR <td>HSSP: P27695; IHD7.</td>	HSSP: P27695; IHD7.
DR <td>InterPro: IPR000097; Apendonc1se1.</td>	InterPro: IPR000097; Apendonc1se1.
DR <td>InterPro: IPR004442; ExoDnase_III.</td>	InterPro: IPR004442; ExoDnase_III.
DR <td>InterPro: IPR004808; ExoIII_xth.</td>	InterPro: IPR004808; ExoIII_xth.
DR <td>InterPro: IPR005135; Exo_endo.phos.</td>	InterPro: IPR005135; Exo_endo.phos.
DR <td>Pfam: PF03372; Exo_endo.phos.1.</td>	Pfam: PF03372; Exo_endo.phos.1.
DR <td>TIGRFAMS: TIGR00195; exoDnase_III.1.</td>	TIGRFAMS: TIGR00195; exoDnase_III.1.
DR <td>TIGRFAMS: TIGR00633; xth.1.</td>	TIGRFAMS: TIGR00633; xth.1.
DR <td>PROSITE: PS00726; AP_NUCLEASE_F1.1; 1.</td>	PROSITE: PS00726; AP_NUCLEASE_F1.1; 1.
DR <td>PROSITE: PS00727; AP_NUCLEASE_F1.2; 1.</td>	PROSITE: PS00727; AP_NUCLEASE_F1.2; 1.
DR <td>PROSITE: PS00728; AP_NUCLEASE_F1.3; 1.</td>	PROSITE: PS00728; AP_NUCLEASE_F1.3; 1.
KW <td>DNA repair; Lyase; Nuclear protein.</td>	DNA repair; Lyase; Nuclear protein.
FT <td>INT_MET 0</td>	INT_MET 0
FT <td>METAL 67 67 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 67 67 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>METAL 95 95 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 95 95 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>METAL 209 209 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 209 209 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>METAL 211 211 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 211 211 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>METAL 307 307 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 307 307 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>METAL 308 308 MAGNESIUM OR MANGANESE (BY SIMILARITY).</td>	METAL 308 308 MAGNESIUM OR MANGANESE (BY SIMILARITY).
FT <td>ACT_SITE 308 308 GENERAL BASE (BY SIMILARITY).</td>	ACT_SITE 308 308 GENERAL BASE (BY SIMILARITY).
FT <td>CONFLICT 20 21 PE -> LP (TN REF. 2).</td>	CONFLICT 20 21 PE -> LP (TN REF. 2).
FT <td>SEQUENCE 317 AA; 35438 MW; 56A7B2029B0891B4 CnC64;</td>	SEQUENCE 317 AA; 35438 MW; 56A7B2029B0891B4 CnC64;
Query Match	4.6%; Score 91; DB 1; Length 317;
Best Local Similarity	21.3%; Pred. No. 1.5;
Matches 51; Conservative 36; Mismatches 74; Indels 78; Gaps 11;	
Oy	83 PROPTSKSEAYVDTLMEDA-----NDTTLIASPSCPTLEDSSPTITWNIGL 134
Db	16 PTTEPEAKSKRAGAKKNKEAEGEVALYEDPDQKTSBGC---KSATLKCSMNVDGL 71
Oy	135 DGCNLPKRA---RGVCSCLALYSPDYVFLDEV-----IPYCAVLLKKRAAYTITGN 184
Db	72 -----RAMIKKKKL-DWVKEEAPDILDLQTKCSKKLPVELDELGSLSHQVMSAPSD 123
Oy	185 EEGYITALLLKGRVKFKFSQELIIPPNTKMMRNLLCVANVSLGNE-----F 230
Db	124 KEYSGVGLSR-----OCPLKVSIGIEEHDGRIYVAEYDAF 164
Oy	231 CLMTSHLEST-----REHSERIROLKTVLKCMQEAPESTVIVFAGDTNLDQEV 280
Db	165 VLVATVAYENAGRGVRLLEYRQKMDAFLR-KFLKGLASRKP---LVLCGLDNLVAHEEL 217
RESULT 12	
CYAA_PROMI	STANDARD: PRT: 865 AA.
AC	Q59685;
DT	01-NOV-1997 (Rel. 35, Created)
DT	01-NOV-1997 (Rel. 35, Last sequence update)
DT	15-DEC-1998 (Rel. 37, Last annotation update)
DE	Adenylate cyclase (EC 4.6.1.1) (ATP pyrophosphate-lyase) (Adenyl1y1
DE	cyase).
GN	CYA
OS	Proteus mirabilis.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
OC	Proteus.
OX	NCBI_TaxID=584;
RN	[1]


```

RP  SEQUENCE FROM N.A.
RX  MEDLINE=97028791; PubMed=8874804;
RA  Trotter P., Sismetro O., Vivas C., Glaser P., Bresson-Roy A.,
RA  Danchin A.;
RT  "Comparative analysis of the cya locus in enterobacteria and related
RT  gram-negative facultative anaerobes.";
RL  Biochimie 78:277-287(1996).
CC  -|- CATALYTIC ACTIVITY: ATP = 3',5'-cyclic AMP + diphosphate.
CC  -|- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -|- SIMILARITY: BELONGS TO ADENYLATE CYCLASE CLASS-1 FAMILY.
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL: U22969; AAC44330.1; -.
DR  InterPro: IPR000274; Adenylt_cyclase_1.
DR  Pfam: PF01295; Adenylate_cyclase_1.
DR  PROSITE: PS01092; ADENYLATE_CYCLASE_1_1; 1.
DR  PROSITE: PS01093; ADENYLATE_CYCLASE_1_2; 1.
KW  Lyase; GMP synthesis.
FT  DOMAIN 1 340 CATALYTIC (POTENTIAL).
FT  DOMAIN 546 865 REGULATORY (POTENTIAL).
SQ  SEQUENCE 865 AA; 99449 MW; 99D4BD51FCB7D7 CRC64;

Query Match      4.6%; Score 90.5; DB 1; Length 865;
Best Local Similarity 21.8%; Pred. No. 7;
Matches 49; Conservative 25; Mismatches 52; Indels 99; Gaps 12;

QY  152 LSPDVLQEVIPPYCA-----YKKRAASTYITGNE-----GYFTAIL 193
    | : : : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  448 LSEPHLTFT--YVPGRANRSGMYLYNRADEPAHIVGHQPLEYNNYLNKLWAMSYFNCL 505
QY  194 LKGRV-----KRSQELI-----PPNTKMM-----RL--LCVNV-- 223
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  506 TKSOVYIHOGSSCDEIKLHEVRDISHPFIRLPAPPRALYSPEIRLAIIVNEV 565
QY  224 -----SLGNEFCIM-----TSLSTREHSAER 247
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  566 DPTERSDQVHDFRKLDFVFSGEDEQLLIDSLYRNSMNEVRYTLHFNGTQ----SM 621
QY  248 IRLKTVLGRKOE--APDSTYVIFAGDTNLR-----DQEVYIK 283
    | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
DB  622 LESLKTITLGMQDAPPAVSVEFCYSOHLRGLIRFVQQLVSEC 666

RESULT 13
YAFD.ECOLI
ID  YAFD.ECOLI STANDARD; PRT; 266 AA.
AC  P30865; P75671;
DT  01-JUL-1993 (rel. 26, Created)
DT  16-OCT-2001 (rel. 40, Last sequence update)
DE  16-OCT-2001 (rel. 40, Last annotation update)
DE  Hypothetical protein yafD.
GN  YAFD OR B0209 OR 20232 OR ECOS0205.
OS  Escherichia coli.
OC  Escherichia coli O157:H7.
OC  Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC  Escherichia.
OX  NCBI_TaxID=562, 83334;
RN  (1)
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12;
RA  Nishimura K., Komine Y., Miyamoto K., Kitabatake M., Mathunaga F.,
RA  Hisano T., Miki T., Inokuchi H.;
RL  Submitted (JUL-1992) to the EMBL/GenBank/DBJ databases.
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / MG1655;

```

```

RX  MEDLINE=97426617; PubMed=9278503;
RA  Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA  Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA  Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA  Mau B., Shao Y.;
RT  "The complete genome sequence of Escherichia coli K-12.";
RL  Science 277:1453-1474(1997).
RN  [3]
RP  SEQUENCE FROM N.A.
RC  STRAIN=K12 / W3110;
RA  Takemoto K., Mori H., Murayama N., Kataoka K., Yano M., Itoh T.,
RA  Yamamoto Y., Inokuchi H., Miki T., Hatada E., Fukuda R., Ichihara S.,
RA  Mizuno T., Makino K., Nakata A., Yura T., Samped G., Mizobuchi K.;
RT  "Systematic sequencing of the Escherichia coli genome: analysis of the
RT  4.0 - 6.0 min (189,987 - 281,416bp) region.";
RL  Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
RN  [4]
RP  SEQUENCE FROM N.A.
RA  Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
RA  Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA  Laskari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
RA  Davis R.W.;
RL  Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN  [5]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / EDL933 / ATCC 700927;
RX  MEDLINE=21074935; PubMed=11206551;
RA  Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA  Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA  Postell G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA  Grobbeck E.J., Davis N.W., Llam A., Dimalanta E.T., Potamianis K.,
RA  Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA  Welch R.A., Blattner F.R.;
RT  "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL  Nature 409:529-533(2001).
RN  [6]
RP  SEQUENCE FROM N.A.
RC  STRAIN=O157:H7 / RIMD 0509952;
RX  MEDLINE=21156231; PubMed=11258796;
RA  Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA  Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA  Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA  Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT  "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT  O157:H7 and genomic comparison with a laboratory strain K-12.";
RL  DNA Res. 8:11-22(2001).
RN  [7]
RP  SEQUENCE OF 47-259 FROM N.A.
RC  STRAIN=K12;
RX  MEDLINE=92137666; PubMed=1663890;
RA  Kajie S.-I., Ideta R., Yamato I., Anraku Y.;
RT  "Molecular cloning and DNA sequence of dhfr, a gene affecting
RT  anaerobic expression of the Escherichia coli hexaneme nitrite
RT  reductase.";
RL  FEMS Microbiol. Lett. 67:205-211(1991).
RN  [8]
RP  This SWISS-PROT entry is copyright. It is produced through a collaboration
RN  between the Swiss Institute of Bioinformatics and the EMBL outstation -
RN  the European Bioinformatics Institute. There are no restrictions on its
RN  use by non-profit institutions as long as its content is in no way
RN  modified and this statement is not removed. Usage by and for commercial
RN  entities requires a license agreement (see http://www.isb-sib.ch/announce/
RN  or send an email to license@isb-sib.ch).
RN  -----
DR  EMBL: D12650; BAA02172.1; ALT_INIT.
DR  EMBL: AE000130; AAC7314.1; -.
DR  EMBL: DB3536; BAA77880.1; ALT_INIT.
DR  EMBL: U70214; AAB08631.1; ALT_INIT.
DR  EMBL: AE005196; AAG54505.1; -.
DR  EMBL: AP002550; BAB35628.1; -.
DR  EMBL: X60739; -. NOT_ANNOTATED_CDS.
DR  PIR: JS0717; JS0717.
DR  EcoGene: EG11650; yafD.

```


CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U43400; AAC54723.1; -
DR InterPro; IPR003498; DNA_pack_C.
DR InterPro; IPR003499; DNA_pack_N.
DR Pfam; PF02499; DNA_pack_C; 1.
DR Pfam; PF02500; DNA_pack_N; 1.
KW DNA packaging.
SO SEQUENCE 663 AA; 76411 MW; 8002D6FA1D166DEC CRC64;

Query Match 4.58; Score 89.5; DB 1; Length 663;
Best Local Similarity 20.28; Pred. No. 5.8;

Matches 74; Conservative 38; Mismatches 125; Indels 129; Gaps 17;

QY 49 CDP-----TWVSVYLRN-DWQTKALSAFELPENOQWPRQPTSFKSEAYV-----96
DB 79 CDPALNCALSKLFTALTEKTEVOYQKHELGTSFYRN-----PMLQIEKFTODFQ 130
QY 97 -----DLTNEADNTTILEASPSGTPLEDSSTISFTWNIDGDCNLPERARGVCSCLA 151
DB 131 RMICGDEPNTNKKERTKLE-----PRQKSLIHILFF-----ISVTKLPTLANHVDLYLK 180
QY 152 LYSDDVVFLOE-----VIIPYCAVLRKRAASYTLITGNEEG 187
DB 181 -YKFDIEFINSSVNLKOKASVFLVPRRHGKTWFMIPVCFILKN-----LEGISTG 232
QY 188 Y-----FTAILLKGVKFKSOEILPEPTKAMRNIL-----CVNV-SL 225
DB 233 YVAHQKHVSHFVKD--VEFKCRFPQKNITGQDNVITIEHETIKSTALFASCYNTHSI 290
QY 226 GGNPECLM--TSHLESTREHSAERIRQLKTYLGKMQEAPDSTTVIFAGDTNLROQEVIR 282
DB 291 RGOSFNLLIYDESHFTKDAFS-----TILGFLPOS--STKILFISTSTNGNHSTSF 340
QY 283 CGGLPDPNVFD-----AMEFLGKPKHCQYTWDTKANNLRI 317
DB 341 LTKLSNPFEMLTVVSVCEDHVIILDRGNATTCACYRIHKPKFTSINADVKKTADLFL 400
QY 318 PAAYKH 323
DB 401 EGAFKH 406

Search completed: November 14, 2002, 10:25:48
Job time : 10.5639 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 ; Search time 13.8671 Seconds

(without alignments)
2565.041 Million cell updates/sec

Title: US-09-697-863a-4

Sequence: 1 MASGSSDAEPGAPAGRAA.....DCGRPPDHGGLCTLNVL 370

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	432.5	22.0	362	T37272	hypothetical prote
2	275	14.0	441	B86252	hypothetical prote
3	118.5	6.0	253	T35506	hypothetical prote
4	111.5	5.7	252	D35853	conserved hypothet
5	105.5	5.4	267	H64044	exodeoxyribonuclea
6	102	5.2	265	F97566	probable nuclease
7	102	5.2	265	AD2787	exonuclease III [i
8	101	5.1	1114	T30299	dynein heavy chain
9	100.5	5.1	502	T40792	hypothetical prote
10	99.5	5.1	556	S32634	splicing factor
11	99.5	5.1	678	H96552	hypothetical prote
12	98	5.0	1495	S27001	alpha-2-macroglobu
13	97	4.9	2218	B84683	hypothetical prote
14	96.5	4.9	840	S74707	nitrogen fixation
15	94.5	4.8	549	T02790	hypothetical prote
16	94	4.8	1240	T48800	SM14 related prote
17	93.5	4.8	282	T51013	related to calmodu
18	93.5	4.8	1246	S56752	helicase SKI2W - h
19	93	4.7	19	H83327	exodeoxyribonuclea
20	93	4.7	317	A39500	DNA (apurinic or a
21	92.5	4.7	288	T24066	exonuclease III ho
22	91	4.6	318	S26830	DNA (apurinic or a
23	90	4.6	266	C64745	yafD protein [simi
24	90	4.6	266	E90654	yafD protein [simi
25	90	4.6	266	E85505	methy-accepting c
26	90	4.6	608	C87638	cell division cycl
27	89.5	4.5	477	S50477	hypothetical prote
28	89.5	4.5	477	S50477	hypothetical prote
29	89.5	4.5	663	T41963	hypothetical prote

30	89.5	4.5	750	2	AB0708	catalase (EC 1.11.
31	89.5	4.5	923	2	A86334	T20H2.17 protein -
32	89	4.5	318	2	S23550	DNA (apurinic or a
33	89	4.5	571	2	H97070	membrane associate
34	88.5	4.5	838	2	G84599	probable retroelem
35	88.5	4.5	1196	2	S35994	DNA repair protein
36	88	4.5	290	2	AC3417	DNA polymerase, ba
37	88	4.5	361	2	S68268	apurinic/apurimid
38	88	4.5	362	2	A56745	glucose-6-phosphat
39	88	4.5	563	2	F64130	microfilin-associ
40	87.5	4.4	536	2	AH3227	genec protein [imp
41	87	4.4	259	2	AH0533	conserved hypothet
42	87	4.4	342	2	AC1266	S-adenosylmethioni
43	87	4.4	497	2	B82981	probable aldehyde
44	87	4.4	875	2	H81739	alanyl-tRNA synthe
45	86.5	4.4	435	2	G84518	Mutator-like trans

ALIGNMENTS

RESULT 1

T27272

hypothetical protein Y63D3A.4 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C:Accession: T27272

R:White, S.

submitted to the EMBL Data Library, October 1998

A:Reference number: Z20335

A:Accession: T27272

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-362 <WIL>

A:Cross-references: EMBL:AU032652; PIDN:CAA21707.1; GSPDB:GND00019; CESP:Y63D3A.4

C:Experimental source: clone Y63D3A

C:Genetics:

A:Gene: CESP:Y63D3A.4

A:Map position: 1

A:Introns: 18/1; 43/1; 116/3; 148/2; 225/3; 320/2

Query Match

Best Local Similarity 22.0%; Score 432.5; DB 2; Length 362;

Matches 113; Conservative 55; Mismatches 126; Indels 43; Gaps 11;

QY	43	FALVGGDPTWPSVLR	ENDMOTQKALSAFELP	ENDQGWPRPPTSFKSEAYVDLT	NED 102
DB	52	FAITATDEAFQSI	LDVDWDLKALDVF	-----YGEAFA-----	E 89
QY	103	ANDTTILEAS	-----PGSTPL	---ED--SSTISFTW	INDIGDCNLPERARGVCSCLATY 153
DB	90	ARSAAVMGASSMS	ASGAAVMTAEDLK	GFVSVMSVNMIDG	RSILTRMKAVAHIVKNV 149
QY	154	SPDVVFQEVTPY	CAVLRKRA	SYTITIGNEE	-GVFTALLKGRKKFKRSQITIEPPNT 212
DB	150	NPDLFLQEVVD	RDLAPIDKLS	IKYYSNKGQCYT	ITALVSK-MFDVEKHDIYHFNQS 208
QY	213	KMMENLLCVN	SLSGNEFCMT	SLHSTREHSA	ERTROLTKTVLGKMQE--APDSTYIV 269
DB	209	GMRYTLLIL	EGSISIGLKV	FLNTHLESTR	REHRRQCAQGFCDKVRREITIAQNPGLVFF 268
QY	270	AGDTNLRDQ	EYIKGGLPD	NVFAWFLGKPKHCQ	YTWDTKANNLRIPAAVKKHREDRIE 329
DB	269	GGDINTLDE	VEYSR--VPD	GVKDAWEA	AGSDNKTFTWDTFKNDKQGFHGAAMRFDRLY 325
QY	330	FAAECHLIP	OSLDVLEK	L-DCGRPP	SDHGLTCT 365
DB	326	W---SGPLD	KVKFTL	EGRQIR	SCLCPPSDHAINAT 359

RESULT 2
B86252
hypothetical protein [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86252
R:TheoLogist, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Hutzar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: B86252
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-441 <STO>
A:Cross-references: GB:AE005172; NID:94835775; PIDN:AMD30241.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match 14.0%; Score 275; DB 2; Length 441;
Best Local Similarity 31.4%; Pred. No. 4.2e-15;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

OY 111 ASPSTPEPDSSTISFTWINDGCLDGCNLPERAGVSCALYSPDVVFLQEVIPPCAY 170
DB 163 ASDSTPLTKILISYNWFERDE--LNRMRIGHILQHSBHLCEFEVPEIYDI 219
OY 171 LKRR-----AASYITINGEGYFTAILKKGRVKEKSOEIIFFPNTKMRNLGVNS 224
DB 220 FPKSMWKAYSCSVSDVAVNGYCMILSLGVKSFSSKS--FGNSIMRELSIAEVE 276
OY 225 LCGNE-FCLMTSHLES-----TREHSAERIRQLKVLKMQAPRSTVIFAGDTNL 275
DB 277 VGRKPLVAFATSHLESPPCGPKMDQFSRBRVQAKAIEILR--PNAVVIKGGDMNW 333
OY 276 RQGEVTKCGG--LPDVFDAWEFLGKPKHCQYWDTKANNLRIIPAAYKRRPFIFFRA 332
DB 334 CD----KLDGKRPPLDKVNDVWEVL-KFGDLAGFTYDTYANMMLSGNRLAKRDLRIICRL 388
OY 333 EBGHLIPQSLDLVG-----LEKLDGGRPSDHWGLLCTLN 367
DB 389 DDYKL--GGEIEMVKEALPGLSYVEKKVKGDKLLEPLVPSDHFGLVTLIS 439

RESULT 3
33506
hypothetical protein SC6E10.19c - Streptomyces coelicolor

C:Species: Streptomyces coelicolor
C>Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35506
R:Seeger, K.; Harris, D.; James, K.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, August 1999
A:Reference number: 221580
A:Accession: T35506
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-253 <SEB>
A:Cross-references: EMBL:AL109661; PIDN:CAB51973.1; GSPDB:GN00070; SCODEB:SC6E10.19c
A:Experimental source: strain A3(2)
C:Genetics:
A:Gene: SCODEB:SC6E10.19c

Query Match 6.0%; Score 118.5; DB 2; Length 253;
Best Local Similarity 23.7%; Pred. No. 0.024;
Matches 68; Conservative 39; Mismatches 115; Indels 65; Gaps 14;

OY 107 TLLEASP-SGTPLDSSSTISFTWINDGCLDGCNLPERAGVSCALYSPDVVFLQEVIP 165
DB TSLPALPNSRTEPDGSAVIRLVINIRSLRD-DTDALARKVIAC---APDLVLLQEA-P 56

OY 166 PYCAVLT-----RAASYITINGEGYFTAILKK-KGRVKEKSOEIIFFPNTKMRNL 219
DB 57 RFRFRKKTTRLAADLVLLSGGAGPALLCSLRATVETEDVLLPLFGRRRRGIA 116
OY 220 CVNVSLGNEFCIMTSHLESTREHSAERIRQLKVLKMQAPRSTVIFAGDTNLROE 279
DB 117 AAVVIRIGARIGVLSHLSLSDAD---ERHEQAGLLDHL-ALGVKRVHVGAGDLNERP-- 170
OY 280 VIKCG-----LPDVFDAWEFLGKPKHCQYWDTKANNLRIIPAAYKRRPFIFFRAE 334
DB 171 ---GGRFRRRIGEGRLDCW--TAAWGEYTFPATAP-----RRIDAVYTE-- 213
OY 335 GHLIPQSLDLVGLEKLDG-----RPSDHWGLLCTLVN 368
DB 214 -----GTEVLCGVPSGLGAVARDLRAATDHLPLVLTALRV 249

RESULT 4
D95853
conserved hypothetical protein [imported] - Sinorhizobium meliloti (strain 1021) mega
C:Species: Sinorhizobium meliloti
C>Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001
C:Accession: D95853
R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Her
Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001
A:Title: The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-fixing e
A:Reference number: A95842; MUID:21396508; PMID:11481431
A:Accession: D95853
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-252 <KUR>
A:Cross-references: GB:AL951985; PIDN:CAC48492.1; PID:g15139964; GSPDB:GN00167
A:Experimental source: strain 1021, megaplasmid pSymB
R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubl
pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.
L.; Hyman, R.W.; Jones, T.
Science 293, 668-672, 2001
A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau
hebaul, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yen,
A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.
A:Reference number: A96039; MUID:21368234; PMID:11474104
A:Contents: annotation
C:Genetics:
A:Gene: SMD20092
A:Genome: plasmid

Query Match 5.7%; Score 111.5; DB 2; Length 252;
Best Local Similarity 21.8%; Pred. No. 0.093;
Matches 62; Conservative 36; Mismatches 102; Indels 85; Gaps 11;

OY 124 ISFTWINDGCLDGCNLPERAGVSCALYSPDVVFLQEV-----IPYCAVL 171
DB 12 IREFTVWVHSCFGDRDLDPARIAVIAEQPRVIALQEDVGARATGGIDQNHMIATHL 71
OY 172 KRAASTYITINGEGYFTAILKKGRVKEKSOEIIFFPNTKMRNLGVNSLGNEFC 231
DB 72 NMEAFHPALHLEDEKGDVLT--ALPRLIKAAPLSSSPRALWEIDVAVKLD 128
OY 232 LMSHLESTREHSAERIRQLKTVLKG--WQAPDSTVTFAGDTN----- 274
DB 129 VIVTHL---RGAERLRQATVALLGPGWLCGMAQGAHVYLAGDLNATGRSTAYRLARQ 185
OY 275 LRQEVIKCGCLPDVNDVDAWEFLGKPKHCQYWDTKANNLRIIPAAYKRRPFIFFRAE 334
DB 186 LSDAQLL-TGVKRRPFPSP-----RLPFL--RIDHV----- 213
OY 335 GHLIPQSLDLV--GLEKLD-----GRPSDHWGLLCTLVN 370
DB 214 -----LVKGIEVASCVRHGSTLARSASDHLPLDLAVDM 249

RESULT 5
H64044

exodeoxyribonuclease III (EC 3.1.11.2) - Haemophilus influenzae (strain Rd KW2)
C.Species: Haemophilus influenzae
C.Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 18-Feb-2000
C.Accession: H64044
R.Fritschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, P.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Wetman, J.; /, D.M.; Brandon, R.C.; Fife, L.D.; Fritchman, J.L.; Fuhrman, J.L.; Geohagen, N.S.M. Science 269, 496-512, 1995
A.Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.Lille: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A.Reference number: A64000; MUID:9550630; PMID:7542800
A.Accession: H64044
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-267 <TIGR>
A.Cross-references: GB:U32689; GB:U42023; NID:g1572980; PIDN:AAC21719.1; PID:g1572987; T
C.Superfamily: exodeoxyribonuclease III
C.Keywords: DNA repair; hydrolase

Query Match	5.4%	Score 105.5	DB 2	Length 267
Best Local Similarity	22.2%	Pred. No. 0.33	Mismatches 63	Conservative 43
			Indels 73	Gaps 14
QY	124	ISFTTNIDGLDCNLPERARGVSCALYSPDVVFLDE--VIPPYCAYLKRAASYTII	181	
		: :	: :	: :
Db	1	MKFTSPININGIRA-----RHOLEAIIKEYQPDVIGLOEIKVADEAPPEYTEITLGYHVF	55	
		:	:	:
QY	182	TGNEGEFTAILLKKGKGVKKSEIIPRPNTMMNMLICVANSIGNPFQMT---	234	
		:	:	:
Db	56	HHGKGHGYALLTKQEPKY-IRGEPTDNEAQRIIMADLE---TEFGLTIVYNGYRP	111	
		:	:	:
QY	235	---SHLESTREHSAERI-RQLKTVLGKMOEADPTSTTVIFAGDTNLDOEYI-----	281	
		: :	: :	: :
Db	112	QGESRAHEKTFPAKEKRYADLOQYLEK--EHRKSNPILIMQDNMISSDLDIGIGDENRK	169	
		:	:	:
QY	282	-----KCGGLPD-----NWFAMEPLGKPKHCQYTW--DTKA-----NNNLRIIP	318	
		:	:	:
Db	170	RWLRTGSCSLPEPRRMYQLYDGLDEFSKRLNLTADKRSWFDYNSKSGEDDNRGLRT--	228	
		:	:	:
QY	319	AAVYKRPDRIFRR---AEBGHLPOSLLVGLTEKIDCGRPSPDH	359	
		:	:	:
Db	229	-----DHILVSOKLAERCVDGIALDIRAEK-----PSDH	259	
		:	:	:

```

RESULT 6
P97566
probable nuclease (PA4172) [Imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002
C:Accession: F97566
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Quirollo, B.; Goldman,
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: F97566
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-265 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK87487.1; PID:g15156812; GSPDB:GM00169
C:Genetics:
A:Gene: AGR_C_3151
A:Map position: circular chromosome
C:Superfamily: exodeoxyribonuclease III

```

```
Query Match Similarity      5.28; Score 102; DB 2; Length 265;
Best Local Similarity      23.58; Pred.No. 0.63;
Matches    67; Conservative   34; Mismatches   94; Indels   90; Gaps   16;
```

```

QY 186 EGYEATLLKKGVKRSKSEI---IPFNTKMRNNLLCVNYSVSGNPFCLMTHSLESTPE 242
Db 60 KGEWGVALIS---KKKPEVNRKGIPLGDADDEQARFIEGVSVBGGKIRCSLYLENGNP 115
QY 243 HS-----AERIRQLKTVLGKMOEADSDTVTFAGDTMLRDQEVKSGGLEPDNWF 291
Db 116 PDDEVKYKPYKLAMELRFRFADRLAME-----PLLIAGDYNV-----IPE-PF 159
QY 292 D---AME---FLGPKK-----HCQY*-----WDTKA-----NNNL 315
Db 160 DCHDDPRWEDDALEFLKPTRAAFRKLENIIGFTDARATTTDAAGLYSEWDYOAGAMPKNNGI 219
QY 316 RIPAAKYKHFREDRIFFRAEGHILPQSLDLYGLEK-LLOGGRPSDH 359
Db 220 RI-----DHMLSAENA---DRLOSVELTEKHVRAWEKPSDH 252

```

RESULT 7
AD2787
exonuclease III [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C.Species: Agrobacterium tumefaciens
C.Date: 11-Jan-2002 #sequence.revision 11-Jan-2002 #text-change 01-Feb-2002
C.Accession: AD2787
R.Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo
erag, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavlin, T.; Levy, R.; Li, M.; McCl
; Kap, P.; Romero, F.; Zhang, S.
Science 294, 2317-2323, 2001
A.Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam
ster, E.W.
A.Title: The genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A.Reference number: AB2577; PMID:11743193
A.Accession: AD2787
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1265 <NUP>
A.Cross-references: GB:AE008688; PIDN:AAU42714.1; PID:g17740152; GSPDB:GN00186
A.Experimental source: strain C58 (Dupont)
C.Genetics:
A.Gene: Atu1714
A.Map position: circular chromosome
C.Superfamily: exodeoxyribonuclease III

Query Match	5.28;	Score 102;	DB 2;	Length 265;
Best Local Similarity	23.58;	Pred. No. 0.63;		
Matches	67;	Conservative	34;	Mismatches 94; Indels 90; Gaps 16;
QY	128	TWNIDSGDGNLEPERARGVCSCLALYSPDYVLEQV--IPYCAYLKKRAASTYITITGNE	185	
DB	5	TWNINGYKA-----RIENLCKMPLKDSPIVCLQIEKSVDBGFPELREIALGHIETHHQ	59	
QY	186	EGYFTALLKKKGVKRSOEI---IPFNTMMRLKLVNNSLGSNGFCMLMSTHLESTRE	242	
DB	60	KGFNGVALLS-----KKRPDEVNNGILPGDDADEQARFICGVFSVGGAIKRCISLYLNGNP	115	
QY	243	HS-----AERIRQLKTLVGLGMDPADPSTTVIPAGDTNLRDQEVIKCGLEPDNVF	291	
DB	116	PDDEVKYYPKLAWMERIRRFPAEDRLAMEE-----PLIAGDYNV-----IPE-PF	159	
QY	292	D-----AME-----ELKKPK-----HCQYT-----MDTKA-----NNUL	315	
DB	160	DCHDRPWEDGALFLKRTYRAFRKLENLGLFDIARATTDAAAGLSYFWDYAGAWMPRNNOI	219	
QY	316	RIPAAVYKHFRRILFFRAEBSGLIPOSLLVGLER-LDGGCFRPSDH	359	
DB	220	RI-----DHLMSAEAA-----DRLOSVELEKIVRAWEKPSDH	252	

RESULT 8
t30299
dmrln heavy chain isotype 7A - sea urchin (Tripneustes gratilla) (fragment)
C:Species: Tripneustes gratilla
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Nov-2000
C:Accession: T30299

R:Gibbons, B.H.; Asai, D.J.; Tang, W.J.; Hays, T.S.; Gibbons, I.R.
 Mol. Biol. Cell 5, 57-70, 1994
 A:Title: Phylogeny and expression of axonemal and cytoplasmic dynein genes in sea urchin
 A:Reference number: Z20814; MUID:94243035; PMID:8186465
 A:Accession: T30299
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-1114 <GIB>
 A:Cross-references: EMBL:U03978; NID:g433384; PID:g433385; PIDN:AAA63592.1
 C:Genetics:
 A:Gene: DYH7A
 C:Superfamily: dynein heavy chain, ciliary

Query Match 5.1%; Score 101; DB 2; Length 1114;
 Best Local Similarity 21.4%; Pred. No. 5.8;
 Matches 53; Conservative 36; Mismatches 87; Indels 72; Gaps 11;

QY 12 PAGPAGAAAPAAQAQAEEDRVKRRRLQCLGFPALVGCDDPTMPVSVLRENDMOTOKALSA 71
 Db 39 PEGPAGTGT-----ETTKDLAKAVAKQCVFNCSDGLD-----YIALGK 78
 QY 72 YEELEPDQGWPRQPTSPKSEAYVDL-----TNEANDTTILEASPSG 115
 Db 79 FFKGLASOGAW-----SCDFENRIDLVLVSVAQQLITLORGINAGA-DTLLEF-----G 128
 QY 116 TPLEDSSTIS-FITWNIDGLDGNLPERARGVCSCLALYSPDVVLEQVIPPYCAVYIKKR 174
 Db 129 TEILDLPTCSFITMNPQYAGRSGLPDLNLKALFRTVAMVPDVALISEIVLYSCGFIKAR 188
 QY 175 AASTYIT-----GNEGY-----FTAILLKGRKFK-----SOELIPPNNKKM 215
 Db 189 PLVKIVATVRYLCSQSLSHHYDYGMRVAKSVLTAAAGNLKLYPEDEDL-----ML 242
 QY 216 RNILCVNV 223
 Db 243 KSIINDVNL 250

RESULT 9
 T40792
 hypothetical protein SPAC9B6.11c - fission yeast (Schizosaccharomyces pombe)
 C:Species: Schizosaccharomyces pombe
 C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jun-2000
 C:Accession: T40792
 R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Volckaert, G.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z21875
 A:Accession: T40792
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-502 <WOO>
 A:Cross-references: EMBL:AL049769; PIDN:CA842372.1; GSPDB:GN00067; SPDB:SPAC9B6.11c
 A:Experimental source: strain 972h-; cosmid c9B6
 C:Genetics:
 A:Gene: SPDB:SPAC9B6.11c
 A:Map position: 2
 A:Introns: 76/1
 C:Superfamily: Schizosaccharomyces pombe hypothetical protein SPAC9B6.11c

Query Match 5.1%; Score 100.5; DB 2; Length 502;
 Best Local Similarity 18.1%; Pred. No. 2.1;
 Matches 88; Conservative 61; Mismatches 154; Indels 183; Gaps 18;
 QY 62 DMOTOKALSAVE-----LPENDQGWPRQPTSPKSEAYVDLTNEDANDTTILEAS 112
 Db 22 DRLSLAISFPDQKKMKKADSIYKKKGSAGKP-SFYTPYIKOKOKLEKAKKAA 80
 QY 113 -----PSGTPLE-----DSSTISFITWNIDGLDGNL----- 139
 Db 81 RKPAPPSNAPPEFNDFIKRMLSLPNVAPETEKSLADITIMTYNV--LAQTNRISM 138
 QY 140 -----PERARGVCSCLALYSPDVVLEQVIPPYCAVYIKKR-----AASITIT--ITGN 184

Db 139 FPHSGEALMKWNRSMLANELTYYSPTLGCQOEVAEEFVNFYKLLGLGLYLHFIKGE 198
 QY 185 BEGYFTAILLKGRKFKFSQELIP-----PNTKMRNLICV-----NNSLGENECIM 233
 Db 199 GRTHGIMIFMKSSLEK-KVODLTITYDDHDLPGRMNTKNGCCVRLERVDPSRGFLA 257
 QY 234 TSHLESTREHSAERIRQ--LKTVLGKMQEAPDSTYVIFAGDTN-----LRDOEV 280
 Db 238 TTHLWHPYGYSTERLRQAILVKEVKNKAQSHPSMPVFIADDFITEPDITFPLALTRPL 317
 QY 281 IKCGILPDNVDPAMEFL-GKPKHCQYTWDTKANN--
 Db 318 SICQATBDIERSMNVYGESELEKNASTKTENDSDEDEECSSSTSSVESTASTP 377
 QY 315 -----LRIPAAYKHRFDRIFF-----RAE 333
 Db 378 KRIILVONDVYVHYRSFYQOEONPVLSYSGVILVPEBNKNTDHAFTNMAHAY 437
 QY 334 EGHIL-----IPQSLDVLGLEKLDGGRFPDHWGL 363
 Db 438 QGHLDYIFVNRDTSLOTPENOVGEGIKALKALRPLPSEKKEAPLE-GRYPSDVALM 496
 QY 364 CTLNVV 369
 Db 497 ANVQIV 502

RESULT 10
 S22634
 sphingomyelin phosphodiesterase (EC 3.1.4.12) precursor - Leptospira interrogans
 N:Alternate names: sphingomyelinase C
 C:Species: Leptospira interrogans
 C:Date: 04-Dec-1992 #sequence_revision 04-Dec-1992 #text_change 08-Oct-1999
 C:Accession: S22634; S10406
 R:Seegers, R.P.A.M.; van der Drift, A.; de Nijis, A.; Corcione, P.; van der Zeijst, B.A.
 Infect. Immun. 58, 2177-2185, 1990
 A:Title: Molecular analysis of a sphingomyelinase C gene from Leptospira interrogans
 A:Reference number: S22634; MUID:90307220; PMID:2163985
 A:Accession: S22634
 A:Molecule type: DNA
 A:Residues: 1-556 <SEG>
 A:Cross-references: EMBL:X52176; NID:g444012; PIDN:CAA36424.1; PID:g444013
 C:Keywords: phosphoric diester hydrolase
 F:1-27/Domain: signal sequence #status predicted <SIG>
 F:28-556/Product: sphingomyelin phosphodiesterase #status predicted <MAT>

Query Match 5.1%; Score 99.5; DB 2; Length 556;
 Best Local Similarity 19.2%; Pred. No. 2.9;
 Matches 60; Conservative 51; Mismatches 106; Indels 95; Gaps 15;

QY 97 DLTNEDANDTTILEASPSGTPLEDS-----STISFITWNI-----DGLDGNLPERANG 145
 Db 55 DLTGSGSVSSPADAPENSILANSIPENMGKILITHNVFLPKTLPGMGNGQNERAQR 114
 QY 146 VCSCLALYSPDVVLEQVLP-----PYCAVILKKRAASYITITGNEGY-- 188
 Db 115 IVSNTYIQNDVIVFDPAPFTDAKTLIDGVRESEPY-----QTDVIGRTKKGMDA 165
 QY 189 -----FT--AILLKGRVYKFSQELIP-----FPYTKMRNLICVNSLGN 228
 Db 166 TLGLRTDAFTNGGVIVSKPIEEKIQHYEKKEGCGADVFSN---KGFAYRIDKNR 221
 QY 229 EFCIMTSHLESTRHSA-----ERIRQLTVLG--KMQEAPDSTYVIFAGDTL-----BD 277
 Db 222 KFHILGTHVOAODSGCNLGVSVKNFNEIRDFIDSKTIPKNEVLIAGDLNVIGSRE 281
 QY 278 QEVTKCGGLPDNVDPAMEFLGKPKHCQYTWDTKANNLIRPAAYKHRFDRIFFRAEGHL 337
 Db 282 YHQMCLILNVNP-----KIYGVF-----FTWDTKTN--ELAAIYKKVE----- 319
 QY 338 IPQSLDVLGLEK 349
 Db 320 -PAYLDIYFVSK 330

RESULT 11

H96552

hypothetical protein F5D21.8 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001

C:Accession: H96552

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

ansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86141; MUID:21016719; PMID:11130712

A:Accession: H96552

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1 678 <STO>

A:Cross-references: GB:AE005173; NID:g10092361; PIDN:AG12770.1; GSPDB:GN00141

C:Genetics:

A:Gene: F5D21.8

A:Map position: 1

C:Superfamily: Arabidopsis thaliana probable ATP-binding cassette protein F12L6.1; ATP-h

Query Match

Best Local Similarity 5.1%; Score 99.5; DB 2; Length 678;

Matches 61; Conservative 46; Mismatches 114; Indels 73; Gaps 11;

34 KRRRLQGLGFAIVGCDPTWVPSVLRNDMOTOKALSAFELPNDQGPPOPTSFKSE 93

Db 80 KRRRLQGLGFAIVGCDPTWVPSVLRNDMOTOKALSAFELPNDQGPPOPTSFKSE 119

QY 94 AYVDLTNDANDTTLEASPSGTPLEDSSTISFTTNWIDLDGCMLEPARAGVCSCLALY 153

Db 120 ---LTREISD-IVETATIDMGLSECSDSRTIGNMHLRISG--GEKKRSLALEVLT 170

QY 154 SPDVFLQEV---IPPYCAV---LKKRAAS-----YTIITGN 184

Db 171 KSLFLFLEDPSSGIDLSAFAFVVOILRNIASSGKTAVSSIHQSGEVPALFDLILLSG 230

QY 185 EGGYFTAILKKGRVFKFSQELIPFPNTKMKRN--LLCVAVSAGNDFCLMTHLESTNE 242

Db 231 ETVYFGE--ASATKFFGEAGFPSPSRNPSDHFLLKCVNSDFDNVTAAVLESRRINDSS 287

QY 243 HSAERIRQLKTVLGRKMEAPDS--TTVI-----FAGDTNLRDQEVIKGGL 286

Db 288 FSLHDLHETNTLPDLDPDIPFAETRTTLVKKFCKSLAASRRARIDELASVIGI 341

RESULT 12

527001

alpha-2-macroglobulin - mouse

C:Species: Mus musculus (house mouse)

C:Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 16-Jul-1999

C:Accession: S27001; A55002

R:van Leuven, F.; Torrekens, S.; Overbergh, L.; Lorent, K.; de Strooper, B.; van den Ber

Eun, J.; Blochem, 210, 319-327, 1992

A:Title: The primary sequence and the subunit structure of mouse alpha-2-macroglobulin,

A:Reference number: S27001; MUID:93076803; PMID:1280217

A:Accession: S27001

A:Molecule type: mRNA

A:Residues: 1-1495 <LEU>

A:Cross-references: EMBL:M93264; NID:g199085; PIDN:AAA3508.1; PID:g199086

R:Umans, L.; Senneels, L.; Hilliker, C.; Stas, L.; Overbergh, L.; De Strooper, B.; Van I

Genomics 22, 519-529, 1994

A:Title: Molecular cloning of the mouse gene coding for alpha-2-macroglobulin and target

A:Reference number: A55002; MUID:95095249; PMID:7528166

A:Accession: A55002

A:Molecule type: DNA

A:Residues: 1-40, 'SGIPEKA', 48-159 <UMA>

A:Cross-references: GB:U06977

A:Note: authors translated the codon TCA for residue 41 as Gln, GGA for residue 42 as

and GCC for residue 47 as Pro

C:Genetics:

A:introns: 29/2; 92/3; 144/1

C:Superfamily: alpha-2-macroglobulin

Query Match

Best Local Similarity 5.0%; Score 98; DB 2; Length 1495;

Matches 66; Conservative 34; Mismatches 106; Indels 82; Gaps 15;

QY 26 AQAEDEVRKRRRLQGLGFAIVG-----CDPTWVPSVLRND---WQD-----KALS 70

Db 1149 SOSQESHVYTKALLVAFALGNKAKRSELLESINKDAVEEDSLIMQRPDGVOKYKALS 1208

QY 71 AYFELPNDQGPPOPTSFSEAVV---DLTNDANDTTLEASPSGTPLEDSSTIS-F 126

Db 1209 FY-----QRPAPSAEVEKTVALLAVLTSSSAPTRDLSS-----DLSTASKI 1252

QY 127 ITW-----NIDGLDGCNLEPARAGVCSCLALYSPDVFLQEVIPPYCAVILKKRAASTYTI 181

Db 1253 VKWISKQONSDD-----GL-----LLTQDTVALQALSKYGSATFTRSQREVLY 1296

QY 182 TGNDEGYFTAILKKGRVFKFSQELI-IP-----FPTKMKRNLLCVNVS 224

Db 1297 TSSSTSTFKTFHVSNGNLLDQEVRLPDLGNTVVTGSGSGCVYIQTSLKYNIL-PVA 1354

QY 225 LGNEFCMLTSHL---ESTREHSAERIRQLKTVLGRKMEAPDSITYI 268

Db 1355 DCKAPFALQVNTLPNLFKAEHDRTQLIRINVSYG--ERPSSNMVY 1399

RESULT 13

B84683

hypothetical protein At2g28300 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cross)

C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001

C:Accession: B84683

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Beaulieu, M.L.; Town, C.D.; Fujii, C.Y

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.B.; Umayam, L.; Tallon,

euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84683

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-2218 <STO>

A:Cross-references: GB:AE002093; NID:g4803953; PIDN:AMD29825.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g28300

A:Map position: 2

Query Match

Best Local Similarity 4.9%; Score 97; DB 2; Length 2218;

Matches 74; Conservative 53; Mismatches 133; Indels 86; Gaps 19;

QY 47 GCGDPTWVPSVLRNDMOTOKALSAFELPNDQGPPOPTSFKSE-ATVDLTNDAND 105

Db 1790 GGGG---LVADVLECSSEPO---LQLPSS-----AEPYISGCTLATLPTEENAD 1835

QY 106 TTLEASPSGTPLEDSSTISFTTNWIDLD-----GCNLEPARAGVCSCLALYSPDV 158

Db 1836 SOLANIEPSSSP-----SVYEKNIEAQDDQDYKTAGCELVSG--GSS-----EPQV- 1880

QY 159 FLQEVIPPYC-----AVLKKRAASTYITITGNEGYFTAIL-LKKGRVFKFSQELI 209

Db 1881 ---HLPPSAEPDGDINHVLKTEKESMSVYVGECTAFPSLSLPTEGNAESQIADTEPF 1936

QY 210 PN-TKMKRNLL-----CVNVSIGN-----EFCMLTSHLESTRHSAERIRQLK 252

Db 1937 TSPTVEKNIKDQEVETGGGLVDSTGSCSPQVQLPSPSPAEPMETGMH-LEETKSE 1995
 QY 253 TLVGGKQGEAP-DSTVIYFAGDTNLRDQEVYKCGGLPDNVFDMWELGKPKHC-----QY 305
 Db 1996 TVTETIOLADIDPSFLIVYQNIIEQDQIEGCG-----DLNVSGSGSTERQIQL 2047
 QY 306 TWDTKANNLRIPAAVKKHFRDIFFRAREGHLIPQSLDVLGKELD 351
 Db 2048 SSSAEPEGHILHEAMNSETV---VTEGSELPSLPTMEDENAD 2090

RESULT 14

S74707 nitrogen fixation positive activator protein - Synecchocystis sp. (strain PCC 6803)

N:Alternate names: protein slr1305

C:Species: Synecchocystis sp.

A:Variety: PCC 6803

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 08-Oct-1999

C:Accession: S74707

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. Kaneo, K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-136, 1996

A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synecchocystis

S:Reference number: S74322; MUID:97061201; PMID:8905231

A:Accession: S74707

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-840 <KAN>

A:Cross-references: EMBL:D90901; GB:AB001339; NID:91651897; PIDN:BA016858.1; PID:dl01759

A:Note: the nucleotide sequence was submitted to the EMBL data library, June 1996

C:Genetics:

A:Gene: nilf

C:Superfamily: response regulator homology

C:Keywords: phosphoprotein

F:9-120/Domain: response regulator homology <RRH>

F:57/Binding site: phosphate (Asp) (covalent) <status predicted

Query Match

Best Local Similarity 4.9%; Score 96.5; DB 2; Length 840;

Matches 55; Conservative 48; Mismatches 96; Indels 89; Gaps 13;

QY 154 SPDVYF-----LQEVIPYCAV-----LKKRASTYITIGNEGYFTAIL- 194
 Db 168 SPDVFITGRNSGCLMDVNPQCFQFGLTKQELVGSNQGYFVWDEQQRGSLRDLLW 227
 QY 195 -KKGVRKFSQELIPFNTKMRNLVCVNSLGGNEF-CLMTSHLESREHSARIQLK 252
 Db 228 QKDGIVKFEKREIEVYDQEKQVETMLSGEPMEFRVDCLLFVMDITERRAE--KQLK 285
 QY 253 TVLGKMOEAPDSTVIYFAGDTNL-----RDQEVYKCGGLPDNVFD 292
 Db 286 -ILSQACEOSPASIVITDVGNIITVYVNPKEISGYSKSAEVLGRNPRILKSGNKTQEDYE 344
 QY 293 A-----W-EFLGKPKHCQYTW----TKANNLRIPAAV-----KHFRD 326
 Db 345 LMKKTILASGRNMGSEFHNRRKNGELYMERASISPISNOQGIIVTHVAVKEDITREKQDAE 404
 QY 327 RIFFRAREGHL-----IPQSLD-----LVGLEKLDGGRF 355
 Db 405 ALFHQAHYDHLTGLPRLIADKRLQDAIESALRQKHIFGLMFLDIDNF 452

RESULT 15

hypothetical protein L549.2 [imported] - Leishmania major (strain Friedlin)

C:Species: Leishmania major

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 19-May-2000

C:Accession: B01455; T02790

R:Myler, P.J.; Audleman, L.; devos, T.; Hixson, G.; Kiser, P.; Lemley, C.; Magness, C.;

Proc. Natl. Acad. Sci. U.S.A. 96, 2902-2906, 1999

A:Title: Leishmania major Friedlin chromosome 1 has an unusual distribution of protein-c

A:Accession: B01455
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-549 <PVL>
 A:Cross-references: GB:AE001274; NID:93264850; PIDN:AA024614.1; PID:92978451; GSPDB:G
 A:Experimental source: strain MHOM/IL/81/Friedlin
 C:Genetics:
 A:Gene: L549.2
 A:Map position: 1
 C:Superfamily: Leishmania major probable membrane protein L549.2
 C:Keywords: transmembrane protein

Query Match

Best Local Similarity 4.8%; Score 94.5; DB 2; Length 549;

Matches 79; Conservative 41; Mismatches 149; Indels 133; Gaps 19;

QY 42 GFALVG-----GCDPTWVPSVIRENDWO-TOKALSAF-----ELPNDGQWRQPTPS 89
 Db 17 GGPVVGIMKRDGTGTRTIRPSIW---DAQIAQSALVKKSSLLAREMQAEKMEAREPEPLK 73
 QY 90 EK-----SEAVYDLTND-----ANDTILEA-----SP 113
 Db 74 RKMVKAQVQSLSRDRMFLMSFNMTDAMGSGHTTPVEAVRVVPEFTRNDECDP 133
 QY 114 SG-----TPLEDSTISFTIWNIDGDCNLPF-RARGVSCSLALYSPDYVFLQEVIPY- 167
 Db 134 MACQVDPSIDKEFPPELT-----PEHRRKHVELLRICYDPDIACINEVKRTFF 182
 QY 168 -----CAYLKKRAASTYII-----TGNEGYFTAILKKGRV 199
 Db 183 NTELWRYVFLGTYGLYQSSRGAQVRALEKGDNAAPSNOGKIAEEDIGNVLLFHKGRF 242
 QY 200 -----KFSQELIPEPN--TKMRNLVCVNSLGGNECLMTSHLESTREHSA 245
 Db 243 FPLMNGRIDGQRFHFAHVGMRDKVTNMTLYVACVQLTAGA-----TAADAVRLHEA 296
 QY 246 ERIQLKTVLGKMOEAPDSTVIYFAGDTNLR--DQEVYKCGGLPDNVFDMWELGKPKHC 303
 Db 297 RQVLTILDALGRDADRSQSHYVCGDLNNQADDEPCVL--LRFRFSTHDLVGGPRWT 354
 QY 304 QY-----TWDTKANNLRIPAAVKKHFRIFFRAREGH 336
 Db 355 AMHYRDARAAYDYITTKNVE-----QAHRSKAYAEIEH 392

Search completed: November 14, 2002, 10:28:20

Job time : 18.8671 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 14, 2002, 10:23:41 : Search time 29.9404 Seconds
(without alignments)
1646.697 Million cell updates/sec

Title: US-09-697-863A-4
Perfect score: 1968
Sequence: 1 MASGSSDPAEPAGPAGRAA.....DCGRPSDHMKLCTINVL 370

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1968	100.0	370	21	AAV56020
2	1258	63.9	362	21	AAV56019
3	1258	63.9	362	22	AAAM39841
4	1258	63.9	372	22	AAAM41627
5	1258	63.9	392	20	AAV03182
6	1250	63.5	369	21	AAAB34403
7	1249	63.5	362	22	AAAB93674
8	1154.5	58.7	306	20	AAV03181
9	961	48.8	263	22	ABG22067
10	647.5	32.9	311	22	ABG22068

11	409	20.8	161	20	AAV11868
12	275	14.0	404	21	AAAG26231
13	275	14.0	404	21	AAAG2551
14	275	14.0	426	21	AAAG26230
15	275	14.0	426	21	AAAG2550
16	275	14.0	437	21	AAAG26229
17	275	14.0	441	21	AAAG2549
18	133.5	6.8	76	20	AAV12175
19	125.5	6.4	449	22	ABBG0974
20	98	5.0	816	22	ABBG0634
21	95	4.8	474	23	ABBP39812
22	93.5	4.8	551	22	ABG20239
23	93.5	4.8	1006	23	AAU98904
24	92.5	4.7	319	16	AAAR80147
25	90	4.6	747	22	ABBT1929
26	90	4.6	1011	22	ABBA4703
27	89.5	4.5	750	22	AAU38392
28	89	4.5	318	18	AAW21746
29	89	4.5	345	21	AAAB56610
30	89	4.5	524	19	AAW48893
31	89	4.5	945	22	ABG06053
32	89	4.5	1284	22	ABG06053
33	89	4.5	1284	22	ABG09636
34	89	4.5	1284	22	ABG10793
35	89	4.5	1284	22	ABG14889
36	89	4.5	1703	22	ABG12526
37	89	4.5	2253	21	AAV44301
38	88.5	4.5	668	21	AAAB28595
39	87.5	4.4	354	22	ABBB1565
40	87.5	4.4	577	22	ABG11645
41	87	4.4	342	23	ABBA4866
42	87	4.4	710	22	ABG20795
43	87	4.4	1557	22	ABG02512
44	87	4.4	1681	22	ABG01407
45	87	4.4	1709	22	ABG02341

ALIGNMENTS

RESULT 1
ID AAV56020
AAV56020 standard; Protein: 370 AA.
AC AAV56020;
DT 15-MAR-2000 (first entry)
DE Mouse CD40 receptor associated protein.
KW Antiartherosclerotic; antiarthritic; neuroprotective; dermatological;
KW immunosuppressive; antinflammatory; immunosuppressive; antiallergic;
KW human; CD40 receptor associated protein; CRAP; cytoplasmic domain;
KW tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;
KW TNF receptor associated factor; TRAF; modulator; signalling pathway;
KW diagnosis; NF-kappaB; Jun; kinase; atherosclerosis; multiple sclerosis;
KW arthritis; systemic lupus erythematosus; graft rejection; allergy;
KW graft versus host disease; autoimmune disease.
XX
XX Mus musculus.
XX
XX PN W09955859-A2.
XX
XX PD 04-NOV-1999.
XX
XX PF 28-APR-1999; 99MO-EP03025.
XX
XX PR 29-APR-1998; 98EP-0201392.
XX
XX (VLA-) VIAAMS INTERUNIVERSITAIR INST BIOTECHNOG.
XX
XX Pye SMC, Remacle JEFJG, Huylebroeck DFE;

DR WPI: 2000-062029/05.
 DR N-PSDB: AA247119.
 XX Novel proteins used to treat inflammatory diseases, NF-kappa related
 PT diseases and for improvement of anti-tumor treatments -
 XX
 XX Claim 3, Page 41-43; 48pp; English.
 XX
 CC This sequence represents the mouse CD40 receptor associated protein
 CC (CRAP). CRAP is a functional protein capable of interacting with the
 CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
 CC factor (TNF) receptor superfamily such as CD30 and TNF receptor 1, where
 CC the protein has no homology to TNF receptor associated factor (TRAF)
 CC proteins. The CD40 binding proteins can be used as modulators of the
 CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
 CC CD40-related, NF-kappa related and/or Jun (kinase)-related diseases,
 CC and for the improvement of anti-tumour diseases. Diseases which may be
 CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
 CC lupus erythematosus, graft rejection, graft versus host disease, allergy,
 CC and autoimmune disease. The proteins can be used to sensitize tumour
 CC cells to anti-tumour treatments and to screen for compounds which
 CC interfere with the interaction of the proteins with other protein
 CC components of the TRAF, CD40 or NF-kappa related pathway.
 CC
 XX Sequence 370 AA:
 SQ
 Query Match 100.0%; Score 1968; DB 21; Length 370;
 Best Local Similarity 100.0%; Pred. No. 1,9e-183;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MASGSSSDAAEPGAPGAPRAASAPAAQAEDVRKRRRLQCLGFALVGGCDPTWVPSVURE 60
 DB 1 MASGSSSDAAEPGAPGAPRAASAPAAQAEDVRKRRRLQCLGFALVGGCDPTWVPSVURE 60
 QY 61 NMQQTAKALSAFELPENQGWPRPPSPKSEAVVDLTNEPDANDTTILEASPSGCTPLED 120
 DB 61 NMQQTAKALSAFELPENQGWPRPPSPKSEAVVDLTNEPDANDTTILEASPSGCTPLED 120
 QY 121 SSTSTFTWNIIDGDCNPERARGVSCALYSPDVVFLQEVIPPCAYLKKRAASYTI 180
 DB 121 SSTSTFTWNIIDGDCNPERARGVSCALYSPDVVFLQEVIPPCAYLKKRAASTI 180
 QY 181 ITGENEGYFTAILLKKGRVFKFSOEIIPPNTKMMRNLLCVNVSIGNEFCLMTSHLEST 240
 DB 181 ITGENEGYFTAILLKKGRVFKFSOEIIPPNTKMMRNLLCVNVSIGNEFCLMTSHLEST 240
 QY 241 REHSARIRIOLKTVLKGMOEAPDSTTVIFAGDTNLRDOEVIKCGSLPNNVDAWEPFLGKP 300
 DB 241 REHSARIRIOLKTVLKGMOEAPDSTTVIFAGDTNLRDOEVIKCGSLPNNVDAWEPFLGKP 300
 QY 301 KHCQYTWDTKANNNLRIIPAAYKHRDRIFRAEGHLLPOSIDLVLGTEKLDGCRFPSPDHW 360
 DB 301 KHCQYTWDTKANNNLRIIPAAYKHRDRIFRAEGHLLPOSIDLVLGTEKLDGCRFPSPDHW 360
 QY 361 GLICTLNVYL 370
 DB 361 GLICTLNVYL 370

RESULT 2
 AAY56019
 ID AAY56019 standard; Protein: 362 AA.
 XX
 AC AAY56019;
 XX
 DT 15-MAR-2000 (first entry)
 XX
 DE Human CD40 receptor associated protein.
 XX
 KW Antiartherosclerotic; antiarthritis; neuroprotective; dermatological;
 KW immunosuppressive; antiinflammatory; immunosuppressive; antiatheric;
 KW human; CD40 receptor associated protein; CRAP; cytoplasmic domain;
 KW tumour necrosis factor; TNF; receptor; superfamily; CD30; homology;
 KW

KW TNF receptor associated factor; TRAF; modulator; signalling pathway;
 KW diagnosis: NF-kappa; Jun; kinase; atherosclerosis; multiple sclerosis;
 KW arthritis; systemic lupus erythematosus; graft rejection; allergy;
 KW graft versus host disease; autoimmune disease.
 XX
 OS Homo sapiens.
 PN WO955859-A2.
 XX
 PD 04-NOV-1999.
 XX
 PF 28-APR-1999; 99WO-EP03025.
 XX
 PR 29-APR-1998; 98BP-0201392.
 XX
 PA (VLAAs) VLAAWS INTERUNIVERSITAIR INST BIOTECHNOG.
 XX
 PI Pyre SMC, Remacle JEFUG, Huylebroeck DFE;
 XX
 DR WPI: 2000-062029/05.
 DR N-PSDB: AA247119.
 XX
 PT Novel proteins used to treat inflammatory diseases, NF-kappa related
 PT diseases and for improvement of anti-tumor treatments -
 XX
 XX Claim 2, Page 39-41; 48pp; English.
 PS
 CC This sequence represents the human CD40 receptor associated protein
 CC (CRAP). CRAP is a functional protein capable of interacting with the
 CC cytoplasmic domain of CD40 and/or other receptors of the tumour necrosis
 CC factor (TNF) receptor superfamily such as CD30 and TNF receptor 1, where
 CC the protein has no homology to TNF receptor associated factor (TRAF)
 CC proteins. The CD40 binding proteins can be used as modulators of the
 CC CD40 signalling pathway, especially to diagnose and treat TRAF-related,
 CC CD40-related, NF-kappa related and/or Jun (kinase)-related diseases,
 CC and for the improvement of anti-tumour diseases. Diseases which may be
 CC treated include atherosclerosis, arthritis, multiple sclerosis, systemic
 CC lupus erythematosus, graft rejection, graft versus host disease, allergy,
 CC and autoimmune disease. The proteins can be used to sensitize tumour
 CC cells to anti-tumour treatments and to screen for compounds which
 CC interfere with the interaction of the proteins with other protein
 CC components of the TRAF, CD40 or NF-kappa related pathway.
 CC
 XX Sequence 362 AA:
 SQ
 Query Match 63.9%; Score 1258; DB 21; Length 362;
 Best Local Similarity 68.4%; Pred. No. 4.2e-114;
 Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;
 QY 24 EAAQAE-EDVRKRRRLQCLGFALVGGCDPTWVPSVURENDMOTOKALSAFELPENQGW 82
 DB 12 EAAEEGEEVEVKKRRLLCVFASFASVASCDAVAQCFLAENDMEMRALNSYEPPEVSAL 71
 QY 83 PROPTSPKSEAVVDLTNEPDANDTTILEASPS-GPLEDSSSTFTWNIIDGDCMLPE 141
 DB 72 ERREPETISPKTYVDLTNEPDSTTSKISPSBDTQDENGSMFLITWNIIDGDLNLSE 131
 QY 142 RARGVSCALYSPDVVFLQEVIPPCAYLKKRAASYTIITGENEGYFTAILLKKGRVFK 201
 DB 132 RARGVSYALYSPDVVFLQEVIPPCAYLKKRSNRYEITGHEGYFTAILMKKSHVK 191
 QY 202 KSGEITIPPNTKMMRNLLCVNVSIGNEFCLMTSHLESTREHSARIRIOLKTVLKGMOEA 261
 DB 192 KSGEITIPPNTKMMRNLLCVNVSIGNEFCLMTSHLESTRGHAERKMQALVMYLRKMOEA 251
 QY 262 PDSTTVIFAGDTNLRDOEVIKCGSLPNNVDAWEPFLGKPKHCQYTWDTKANNNLRIIPAAY 321
 DB 252 PSATVYIFAGDTNLRDREVTGCGGLPNNIVDWERFLGPKHCQYTWDTQWNSNLGITAAC 311
 QY 322 KHRPDRIFR--AEGRHLLPOSIDLVLGTEKLDGCRFPSPDHGGLCTLNVYL 370
 DB 312 KHRPDRIFR--AEGRHLLPOSIDLVLGTEKLDGCRFPSPDHGGLCTLNVYL 362

```
RESULT 3
AAM39841
ID AAM39841 standard; Protein; 362 AA.
XX
AC AAM39841;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 2986.
XX
KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AAI58997.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 4; SEQ ID NO 2986; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cyostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 362 AA.
XX
Query Match 63.9%; Score 1258; DB 22; Length 362;
Best Local Similarity 68.4%; Pred. No. 4,2e-114;
Matches 240; Conservative 43; Mismatches 64; Indels 4; Gaps 3;
QY 24 EAAQAE-EDRVKRRRLQCLGFALVGGCDPTWPSVLRENDMOTOKALSAFELPENDGWM 82
```

```
DB 12 EAAEEGEPVEVAKRRLLCYEFASVASCDAAVAOCPFLAENDMEMERALSYFEPVEESAL 71
QY 83 PRQPPISFKSEAVYDLTNDAADTTILEASPS-GPPLEDSSITSTTWNIDGLDCNIP 141
DB 72 ERRPETISEPKTYVDLTNETDSTTSKISPEDTQOENGSMFSLITWNIDGLDNLNISE 131
QY 142 RARGVSCALATSPDVVFOEVIPYCAVILKKRAASYTTITNGEGYTTAILKGRVKF 201
DB 132 RARGVSYALATSPDVVFOEVIPYCAVILKKRASSNYELITGHEGTYTALMLKSRVKL 191
QY 202 KQOEIIPEPNTMMRNLLCVNVLGNEFCMTSHLESTREHSARIRQLKTVLGKMOEA 261
DB 192 KQOEIIPEPSTKMMRNLLCVHNVSGNELCLMTSHLESTRGHAERMNQLKMYLKKMOEA 251
QY 262 PSTVIVIFAGDTNLRDQVYKCGSLPDVVPDAWEFLGKPKHCQYTWDTQKANNRIIPAY 321
DB 252 PESATVIFAGDTNLRDQVYKCGSLPDVVPDAWEFLGKPKHCQYTWDTQKANNRIIPAY 311
QY 322 KHRPDRIFER--AEEGHLIPQSLDVLGLEKLDGCRFPDSHMGLLCTLVNVL 370
DB 312 KLRPDRIFERAAEEGHIIIPRSLDLGLEKLDGCRFPDSHMGLLCNDIIL 362
```

```
RESULT 4
AAM41627
ID AAM41627 standard; Protein; 372 AA.
XX
AC AAM41627;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polypeptide SEQ ID NO 6558.
XX
KW Human; nootropic; immunosuppressant; cyostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokine; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia.
XX
OS Homo sapiens.
XX
PI Wang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI: 2001-442253/47.
DR N-PSDB: AAI60783.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Example 2; SEQ ID NO 6558; 10078pp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
```



```

XX WPI: 2000-587534/55.
DR N-PSDB; AAC98160.
XX Colon cancer associated gene sequences, referred to as colon cancer
PT antigens, useful for the treatment, prevention, and diagnosis of colon
PT disorders such as colon cancer -
XX
PS Claim 11: Page 1508-1509; 2104bp; English.
XX
CC AAC97991 to AAC98763 encode the human colon cancer associated proteins,
CC called human colon cancer antigens, given in AAB53234 to AAB54006. The
CC human colon cancer antigens can have cytostatic, cardioactive, muscular;
CC neuroprotective, immunomodulatory, gynaecological, gastrointestinal,
CC can be used in gene therapy. The colon cancer antigen polynucleotides,
CC proteins and antibodies to the proteins are useful for the prevention,
CC treatment and diagnosis of colon disorders, such as colon cancer. The
CC polynucleotides may be used in diagnostics and research, such as for
CC chromosome identification, and as hybridisation probes. The proteins
CC may also be used to prevent diseases such as neural disorders, immune
CC system disorders, muscular disorders, reproductive disorders,
CC gastrointestinal disorders, wounds, renal disorders, infectious
CC diseases, and cardiovascular disorders. AAC98764 to AAC98772 and
CC AAB54007 represent sequences used in the exemplification of the present
CC invention.
XX
SO Sequence 369 AA:
XX
Query Match 63.5%; Score 1250; DB 21; Length 369;
Best local Similarity 68.1%; Pred. No. 2.6e-113;
Matches 239; Conservative 43; Mismatches 65; Indels 4; Gaps 3;
XX
QY 24 EAAQAE-EDRVKRRRLQCGFALVGGCDPTWVPSVLRNDMOTOKALSAFELPENDQGW 82
DB 19 EAAEEGEPPEVKRRRLQCFEASVASCDAVAOCCFLAENDWEMERALNSYFEPVEESAL 78
QY 83 PROPTSFSEAYVDLTNEDANDTTLLEASPS-GTPLEDSSTFSFTTWIDGDCGNLPE 141
DB 79 ERREPTISEKTYVDLTNETTDTSTTSKISPSEDTQOENGSMSFLTWTWIDGLNNLSE 138
QY 142 RANGVSCSLALYSPDVVFLQEVIPPCAYLKKRAASTYITIGNEGYFTAILLKGRVRF 201
DB 139 RANGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNVEITITGHEEGFTAILMKSRVYL 198
QY 202 KSGEITFPPTKMMRNLLCVNLSIGNEFCLMTSHLESTRHSAERIRQLKTVLGMQEA 261
DB 199 KSGEITFPPTKMMRNLLCVHNVWSGNECLMTSHLESTRGHAERMNQJLKVLMKMOEA 258
QY 262 PDSTTVIFAGDTNLRDQEVYKCGGLPDNVFDMAEFLGKFKHCQYTWDTKANNLRLIPAY 321
DB 259 PESATVIFAGDTNLRDREYTRCGGLPNNIVDWMEFLGKFKHCQYTWDTOMNSNLGITTAAC 318
QY 322 KHFEDLIFRR--AEBGHLIPQSLDVGLEKLDCCGRPPSDHWGLCTLNVYL 370
DB 319 KLFEDLIFRRAAEHGHIIPRSIDLGLEKLDCCGRPPSDHWGLCTLNVYL 369

```

```

XX 07-FEB-2001.
PD 28-JUL-2000; 2000EP-0116126.
XX
XX 29-JUL-1999; 99JP-0248036.
XX 27-AUG-1999; 99JP-0300253.
XX 11-JAN-2000; 2000JP-0118776.
XX 02-MAY-2000; 2000JP-0183767.
XX 09-JUN-2000; 2000JP-0241899.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
DR WPI, 2001-316749/34.
XX
PS Claim 8; SEQ ID 13210; 2537bp + CD ROM; English.
XX
CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences. AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
SO Sequence 362 AA:
XX
Query Match 63.5%; Score 1249; DB 22; Length 362;
Best local Similarity 68.1%; Pred. No. 3.1e-113;
Matches 239; Conservative 43; Mismatches 65; Indels 4; Gaps 3;
XX
QY 24 EAAQAE-EDRVKRRRLQCGFALVGGCDPTWVPSVLRNDMOTOKALSAFELPENDQGW 82
DB 12 EAAEEGEPPEVKRRRLQCFEASVASCDAVAOCCFLAENDWEMERALNSCFEPVEESAL 71
QY 83 PROPTSFSEAYVDLTNEDANDTTLLEASPS-GTPLEDSSTFSFTTWIDGDCGNLPE 141
DB 72 ERREPTISEKTYVDLTNETTDTSTTSKISPSEDTQOENGSMSFLTWTWIDGLNNLSE 131
QY 142 RANGVSCSLALYSPDVVFLQEVIPPCAYLKKRAASTYITIGNEGYFTAILLKGRVRF 201
DB 132 RANGVCSYLALYSPDVIFLQEVIPPYSYLKKRSSNVEITITGHEEGFTAILMKSRVYL 191
QY 202 KSGEITFPPTKMMRNLLCVNLSIGNEFCLMTSHLESTRHSAERIRQLKTVLGMQEA 261
DB 192 KSGEITFPPTKMMRNLLCVHNVWSGNECLMTSHLESTRGHAERMNQJLKVLMKMOEA 251
QY 262 PDSTTVIFAGDTNLRDQEVYKCGGLPDNVFDMAEFLGKFKHCQYTWDTKANNLRLIPAY 321
DB 252 PESATVIFAGDTNLRDREYTRCGGLPNNIVDWMEFLGKFKHCQYTWDTOMNSNLGITTAAC 311

```


CC products. The proteins obtained can be used to develop products for diagnosis and
CC therapy. The proteins obtained may have cytokine activity, cell
CC proliferation and differentiation activity, haematopoiesis regulating
CC activity, tissue growth regulating activity, reproductive hormone
CC regulating activity, chemotactic/chemokinetic activity, haemostatic and
CC thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC activity, tumour inhibition activity or other activities. The products
CC can be used in forensic, gene therapy and chromosome mapping procedures.
CC The sequences can also be used for obtaining corresponding promoter
CC sequences. The nucleic acids encoding the signal peptides can be used for
CC directing extracellular secretion of a polypeptide or the insertion of a
CC polypeptide into a membrane, or importing a polypeptide into a cell.

XX Sequence 161 AA:

Query Match 20.8%; Score 409; DB 20; Length 161;
Best Local Similarity 56.0%; Pred. No. 1.1e-31;
Matches 84; Conservative 19; Mismatches 45; Indels 2; Gaps 2;

QY 24 EAQAQAE-EDRYKRRRLQCLFALYGGCDPTWFSVLRENWQOKALSAFELPEPDQGW 82
DQ 12 EAEEEEPEVEKKRRLLCVFASVASCDAVAOCFLAENDMEMERALNRYFEPPVESAL 71
QY 83 PROPTFSKSPAYVDLTFNEANDTTIEASPS-GTPLEDSSTISFTIWNIDGDCNLP 141
DQ 72 ERREPTISEPTTYDLNNEETDSTSTKISPSSEDTQOENGSMFLITWNIDGLNLS 131
QY 142 RARGVCSCLALYSPDVVFLQEVIRPYCAVL 171
DQ 132 RARGVCSYALYSPDVIFLQEVIRPYSYL 161

RESULT 12

ID AAG26231 standard; Protein: 404 AA.

XX AAG26231:

XX 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 30610.

XX Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.

XX Arabidopsis thaliana.

OS EP1033405-A2.

PN 06-SEP-2000.

PD 25-FEB-2000; 2000EP-0301439.

PF 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123548.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 08-APR-1999; 99US-0128234.

PR 16-APR-1999; 99US-0128714.

PR 19-APR-1999; 99US-0129845.

PR 21-APR-1999; 99US-0130077.

PR 23-APR-1999; 99US-0130449.

PR 28-APR-1999; 99US-0130510.

PR 30-APR-1999; 99US-0131449.

PR 04-MAY-1999; 99US-0132407.

PR 05-MAY-1999; 99US-0132484.

PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-01353629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0138119.
PR 16-JUN-1999; 99US-0138452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140695.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 29-JUN-1999; 99US-0141287.
PR 30-JUN-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 01-JUL-1999; 99US-0142055.
PR 02-JUL-1999; 99US-0142390.
PR 06-JUL-1999; 99US-0142803.
PR 08-JUL-1999; 99US-0142920.
PR 09-JUL-1999; 99US-0142977.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 14-JUL-1999; 99US-0144005.
PR 15-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144332.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 20-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 21-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151348.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.

PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 275; DB 21; Length 404;
Best Local Similarity 31.4%; Pred. No. 5; Le-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASPSGTPLEDSSSTISPTMNDIGLCNLPERRRGVCSCLALYSPDYVLOEVIPYCAV 170
DB 126 ASDSGTPCTCKILSYNVWFREDLE---LNLKRAIGHLIQLHSPHLICFQEVTPETIYDI 182
QY 171 LKKR-----AASYTIIIGNEGYFTAILDKGRVKEKSOETIPPTKMRNLLCVNVS 224
DB 183 FRKSNMMAKAYSCSVSDVAVSRGYCMLSKLGVSFSSKS---FGNSIMGRELSIAEVE 239
QY 225 LGGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGMQADPSTVIFAGDTNL 275
DB 240 VPKRKPLVATSHLESPPCGPKKIDOMFSREKVEQAKAITEILR---PNAVIVIGSGMNV 296
QY 276 RDEVIKCGG---LPDNEFDANEFLGKPKHCQYIWDTKANNLRIPAAVYKHREDRIFFRA 332
DB 297 CD---KLDGKPELPDKWVMEVL-KPGDLGFTYDTKAMPMLGNNRLOKRDRIICRL 351
QY 333 EECGLIPQSLIDLVG-----LEKIDCGRPSPDHMGGLCTLN 367
DB 352 DDYKL--GGIEHWGKAIRIGLSYKEKKYRGDIKKELPVLPDSHFGILVTLIS 402

RESULT 13
AAG42551
ID AAG42551 standard; Protein; 404 AA.
XX
AC AAG42551;
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 53080.
XX
KM Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 14-MAY-1999; 99US-0134768.
PR 18-MAY-1999; 99US-0134941.
PR 19-MAY-1999; 99US-0135124.
PR 20-MAY-1999; 99US-0135353.
PR 21-MAY-1999; 99US-0135629.
PR 24-MAY-1999; 99US-0136021.
PR 25-MAY-1999; 99US-0136392.
PR 27-MAY-1999; 99US-0136782.
PR 28-MAY-1999; 99US-0137222.
PR 01-JUN-1999; 99US-0137528.
PR 03-JUN-1999; 99US-0137502.
PR 04-JUN-1999; 99US-0137724.
PR 07-JUN-1999; 99US-0138094.
PR 08-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 10-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 16-JUN-1999; 99US-0139454.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139761.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139899.
PR 22-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 23-JUN-1999; 99US-0140699.
PR 24-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144003.
PR 16-JUL-1999; 99US-0144086.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 11-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 22-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151338.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.

PR 14-OCT-1999; 9905-0159330.
PR 14-OCT-1999; 9905-0159331.
PR 14-OCT-1999; 9905-0159637.
PR 14-OCT-1999; 9905-0159638.
PR 18-OCT-1999; 9905-0159584.
PR 21-OCT-1999; 9905-0160741.
PR 21-OCT-1999; 9905-0160767.
PR 21-OCT-1999; 9905-0160770.
PR 21-OCT-1999; 9905-0160776.
PR 21-OCT-1999; 9905-0160814.
PR 21-OCT-1999; 9905-0160815.
PR 22-OCT-1999; 9905-0160980.
PR 22-OCT-1999; 9905-0160981.
PR 22-OCT-1999; 9905-0160989.
PR 25-OCT-1999; 9905-0161404.
PR 25-OCT-1999; 9905-0161405.
PR 25-OCT-1999; 9905-0161406.
PR 26-OCT-1999; 9905-0161359.
PR 26-OCT-1999; 9905-0161360.
PR 26-OCT-1999; 9905-0161361.
PR 28-OCT-1999; 9905-0161920.
PR 28-OCT-1999; 9905-0161992.
PR 28-OCT-1999; 9905-0161993.
PR 29-OCT-1999; 9905-0162142.

Query Match 14.0%; Score 275; DB 21; Length 404;
Best Local Similarity 31.4%; Pred. No. 5,1e-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

QY 111 ASFGSPLEDSSTIFSTININDGCMLEFARGVSCSLADVDVDFIOEVIPIPCAY 170
DB 126 ASDSGPLICLKLTSTNVWFREDLE---LNLFRMRAIGHLIOLHSPHLICQEVTPPEIYDT 182
QY 171 LKRR-----AASYTIITNEGEYFTAILLKGRVKEFSOEIIPFPITKMMRLLCVNS 224
DB 183 FRKSNMKAIVSCSVYDVAVSRGYCMILSKLGVKFSFSSKS---FGNSIMGRELSIAEVE 239
QY 225 LGGNE-FCLMTSHLES-----TRESAEBIRQLKTVLGMGEADPSTTVIFAGDTNL 275
DB 240 VPGKRPVATSHLESPPCGPPKWDQMSREKVEQAKAEILRL--PNAVYIFGGDMW 296
QY 276 RDOEVIKCGG---LPDNDVANEFLGKPRHCQYTWDTKANNLRIIPAYKHREDFRIFFRA 332
DB 297 CD-----KLDSKFPPLPKWVDWEVL-KPDGLGFTYDTKANPMLSGNRALQKRDRIICRL 351
QY 333 EEEGLIPOSIDIVG-----LEKLDGGRPPSHMGLICTLN 367
DB 352 DDTKL--GGIEMVKEAIPGLSYVKEKKVGDIKLELPVLPDSHFGLVTL 402

RESULT 14
AAG26230
ID AAG26230 standard; Protein; 426 AA.
XX
AC AAG26230;
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 30609.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EPI033405-A2.
XX
PD 06-SEP-2000.
XX
PE 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 9905-0121825.

PR 05-MAR-1999; 9905-0123180.
PR 09-MAR-1999; 9905-0123548.
PR 23-MAR-1999; 9905-0125788.
PR 25-MAR-1999; 9905-0126264.
PR 29-MAR-1999; 9905-0126785.
PR 01-APR-1999; 9905-0127462.
PR 06-APR-1999; 9905-0128234.
PR 08-APR-1999; 9905-0128714.
PR 16-APR-1999; 9905-0129845.
PR 19-APR-1999; 9905-0130077.
PR 21-APR-1999; 9905-0130449.
PR 23-APR-1999; 9905-0130510.
PR 23-APR-1999; 9905-0130891.
PR 28-APR-1999; 9905-0131449.
PR 30-APR-1999; 9905-0132048.
PR 30-APR-1999; 9905-0132407.
PR 04-MAY-1999; 9905-0132484.
PR 05-MAY-1999; 9905-0132485.
PR 06-MAY-1999; 9905-0132486.
PR 06-MAY-1999; 9905-0132487.
PR 07-MAY-1999; 9905-0132863.
PR 11-MAY-1999; 9905-0134256.
PR 14-MAY-1999; 9905-0134218.
PR 14-MAY-1999; 9905-0134219.
PR 14-MAY-1999; 9905-0134221.
PR 14-MAY-1999; 9905-0134370.
PR 18-MAY-1999; 9905-0134768.
PR 19-MAY-1999; 9905-0134941.
PR 20-MAY-1999; 9905-0135124.
PR 21-MAY-1999; 9905-0135353.
PR 24-MAY-1999; 9905-0135629.
PR 25-MAY-1999; 9905-0136021.
PR 27-MAY-1999; 9905-0136392.
PR 28-MAY-1999; 9905-0136782.
PR 01-JUN-1999; 9905-0137222.
PR 03-JUN-1999; 9905-0137528.
PR 04-JUN-1999; 9905-0137724.
PR 07-JUN-1999; 9905-0138094.
PR 08-JUN-1999; 9905-0138540.
PR 10-JUN-1999; 9905-0138847.
PR 10-JUN-1999; 9905-0139119.
PR 14-JUN-1999; 9905-0139452.
PR 16-JUN-1999; 9905-0139453.
PR 17-JUN-1999; 9905-0139492.
PR 18-JUN-1999; 9905-0139454.
PR 18-JUN-1999; 9905-0139455.
PR 18-JUN-1999; 9905-0139456.
PR 18-JUN-1999; 9905-0139457.
PR 18-JUN-1999; 9905-0139458.
PR 18-JUN-1999; 9905-0139459.
PR 18-JUN-1999; 9905-0139460.
PR 18-JUN-1999; 9905-0139461.
PR 18-JUN-1999; 9905-0139462.
PR 18-JUN-1999; 9905-0139463.
PR 18-JUN-1999; 9905-0139750.
PR 18-JUN-1999; 9905-0139753.
PR 21-JUN-1999; 9905-0139817.
PR 22-JUN-1999; 9905-0139899.
PR 23-JUN-1999; 9905-0140353.
PR 23-JUN-1999; 9905-0140354.
PR 24-JUN-1999; 9905-0140695.
PR 28-JUN-1999; 9905-0140823.
PR 29-JUN-1999; 9905-0140991.
PR 30-JUN-1999; 9905-0141287.
PR 01-JUL-1999; 9905-0141842.
PR 01-JUL-1999; 9905-0142154.
PR 02-JUL-1999; 9905-0142055.
PR 06-JUL-1999; 9905-0142390.
PR 08-JUL-1999; 9905-0142803.
PR 09-JUL-1999; 9905-0142820.
PR 12-JUL-1999; 9905-0142977.
PR 13-JUL-1999; 9905-0143542.

[illegible]

PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158028.
PR	08-OCT-1999;	99US-015823.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159293.
PR	13-OCT-1999;	99US-0159294.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160768.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161992.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

Query Match	14.0%	Score 275	DB 21	Length 426
Best Local Similarity	31.4%	Pred. No. 5	Se-18	
Matches 92	Conservative 38	Mismatches 11	Indels 52	Gaps 11
QY 111	ASPSGTPLEDSSIRISFTNNIDGLDCGNLPERARGSCSLATYSPDVFLQEVIPYCAV	170		
DB 148	ASDSGTPFLCTKLSTNNWFREDLE---LNLRRALGHLIQHLSPLHLCQGEVTPETIYD	204		
QY 171	LKKR-----ASVYTIITGNEGYFTAILLKRGVYFKSOELIIPPNTKMMRNLLCVNS	224		
DB 205	FRKSNMWMKAVSCSVDAVAVSRGYCMLSLKLTGVKSFSSKS---FENSINGRELSTAEVE	261		
QY 225	LGNE--FCLMTSHLES-----TREHSAERITROLKTYGKMQEAPDSTTVIFAGDTNL	275		
DB 262	VPRKPLVFATSHLESPPGPRKMDQFSEKREYQAKAEATIELR---PNANVTFGGDMNV	318		
QY 276	RQDEVIRCGS---LPDVFAMEFLGKPKCOQTYTWDTKANNLNLRIPAAVKKRPDRIFPRA	332		
DB 319	CD----KLDSKFLPLDKMNVYWEVL--KRGDLGTYTDFTKAPMLSGNRALQKRIDRILICRL	373		
QY 333	EEGHLIPQSLDLVG-----LEKLDCGAFPSDDHGMGLCTLN	367		
DB 374	DDVKLL--GGIEVMGKEALPELSTYVKEKKVGDIIKKLELPLPDPDHGELVLTLS	424		
RESULT 15				
AAG42550				
ID	AAG42550 standard; Protein: 426 AA.			
XX				
AC	AAG42550;			
XX				
DT	18-OCT-2000 (first entry)			
DE	Arabidopsis thaliana protein fragment SEQ ID NO: 53079.			
XX				
KW	Protein identification: signal transduction pathway; metabolic pathway;			
KW	hybridisation assay; genetic mapping; gene expression control; promoter;			
KW	termination sequence.			
XX				

PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 13-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160960.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 14.0%; Score 275; DB 21; Length 426;
Best Local Similarity 31.4%; Pred. No. 5.5e-18;
Matches 92; Conservative 38; Mismatches 111; Indels 52; Gaps 11;

OY 111 ASPSGTPLEDSTIFITWNIDGDLNLPFRAGVCSCLALYSPDYVFLQEVIPPCAY 170
DB 148 ASDSGTPLTCLIKILSYNWFREDE--LNLKMRALGHLIQHSPHLICFOEVTPEIYDI 204
OY 171 LKKR-----ASYYTITGNEGEYFTAILLKGRVYFKSDEIIPPTKMRNLGVNS 224
DB 205 FRKSMWMAKAYSCSVADVAVSNGYICMLLSKIGVKSFSKS--FGNSIMGRELIAEVE 261
OY 225 LCGNE-FCLMTSHLES-----TREHSAERIRQLKTVLGMQDEAPDSTIYIFAGDTNL 275
DB 262 VPGRRPLVATSHLESPPCGPPKMDQMSRERVEQAKAEIILR--PNANYIEGDMNW 318
OY 276 RQEVYIKCG---LPDNYVDAMEFLGKPRHCQYTWDTKANNLRIIPAAYKHFEDRIFFRA 332
DB 319 CD---KIDGKPLPLPKWVWEVL-KRGDLGFTYDTKANPMLSGNRALQKRDLRLICRL 373
OY 333 ECHGLIPSLDLVG-----LEKIDCGRFPSPDHMLGLCTLN 367
DB 374 DDKYL--GGIEWVGKAIPLGLSYVKEKKVRGDIKKLELPLPSDHFGLLVTLIS 424

Search completed: November 14, 2002, 10:27:25
Job time : 30.9404 secs